Psychological Bulletin

CONTENTS

[illegible] No. 4

The *Psychological Bulletin* contains evaluative digests of research literature and articles on research methodology in psychology. This JOURNAL does not publish reports of original research or original theoretical articles.

Manuscripts should be sent to the Editor-elect, Harry Helson, Department of Psychology, University of Texas, Austin, Texas.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the "Publication Manual of the American Psychological Association" (1957 Revision). Special attention should be given to the section on the preparation of the references (pp. 59-63), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted *in duplicate.* Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

Reprints. Fifty free offprints are given to contributors of articles and notes. Authors of early publication articles receive no gratis offprints.

Communications—including subscriptions, notices of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee extra copies forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). *Single copies,* $1.50.

PUBLISHED BIMONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Washington, D. C.
and New Era Printing Co., Lancaster, Pa.

Entered as second class matter at the post office at Washington, D.C., under the act of March 3, 1879. Acceptance for mailing at special rate of postage provided for in section 1103, act of October 3, 1917, authorized January 20, 1925.

VOL. 55, NO. 4 JULY, 1958

Psychological Bulletin

PROBLEMS AND METHODS OF PSYCHOPHYSICS[1]

S. S. STEVENS
Harvard University

The methods and procedures of psychophysics have been reviewed from time to time and marshaled into more-or-less logical array. If another such inventory is now in order it is because recent developments allow us to put the various methods in new perspective and to see more clearly how they articulate with the problems of psychophysics.

Let us admit first off that a concern with method is justified only if it leads to something beyond itself. The study of method, which I suppose is the proper meaning of that overworked term *methodology*, is one of those "necessary evils" whose justification lies in its potential contribution to the solving of substantive problems. But the problems are the main concern. If an empirical problem is worth solving, a method for it is worth developing, but it may turn out that there is little profit in fashioning tools to do what nobody wants done. Methodology can easily become methodolatry.

Psychophysical methods have at times been treated as though they were ends in themselves, and in many texts the term psychophysics has seemed to be synonymous with three "classical" procedures for solving an issue that few people care about. Little wonder then that psychophysics has sometimes been accused of inconsequence. Attitudes of this sort are not improved by the decision of a distinguished committee to define psychophysics as the use of a human observer as a "null instrument" to determine "equality or difference of sensations" (**20**, p. 59). According to the view of this committee, psychophysics is a strange land lying between "the physical and psychical" (**20**, p. 65). As Evans (**9**, p. 5) puts it, "Psychophysics at present, therefore, is limited to the *relative* evaluation of light beams with respect to normal observers under standardized conditions. . . ."

Psychophysics is really a much more nutritious subject than these conceptions imply. Seeking the laws that relate the responses of men and animals to the energetic configurations of the environment, it probes matters of deep human interest, and matters that often make a practical difference in the market place. For some of us, at any rate, a certain excitement attaches to the discovery that on "quantitative" or prothetic perceptual continua, such as brightness, loudness, heaviness, length, duration, etc., equal stimulus ratios produce equal sensation ratios (**31**). This principle means that the psychological magnitude is a power function

[1] Supported by a grant from the National Science Foundation and by Contract Nonr-1866(15) with the Office of Naval Research (Project Nr142-201, Report PNR-190). Reproduction for any purpose of the U. S. Government is permitted.

of the physical magnitude. And an example of practical utility can be seen in the application of this power law to the problem of predicting the loudness of a complex noise from physical measurements made on the spectrum of the sound (**29**, **33**).

Psychophysics has its problems and its methods. The purpose of this paper is to try to classify the methods in terms of the problems. It is a follow-up on some earlier attempts (**25**, **26**), in which a similar point of view was tried out in part, but in which the coverage was less systematic. A related effort, but rather different in outcome, was made by Guilford (**14**). In attempting an exercise of this sort we must realize that an element of arbitrariness attaches to taxonomies, and alternative schemes are always possible. Furthermore, we must forego the ambition to be exhaustive and completely consistent, for the methods of psychophysics are in a state of flux, and, as knowledge continues to expand, our understanding of procedures will change and improve.

The names that have become attached to the various methods show interesting vagaries, for in labeling our methods we fasten attention sometimes on one feature and sometimes on another. The name refers sometimes to the manner of presenting stimuli, sometimes to the task assigned the observer, and sometimes to the statistical treatment used to process the data.

Having names for methods provides a convenient shorthand for the description of experiments—as well as some handy items to ask about on examinations. But labeling is not without its drawbacks. Labeling produces jargon, and jargon leads to esoteric discourse. Many readers would find clarity improved if special names for procedures were banned and authors were forced to frame their descriptions in conventional English.

But names for the methods are probably here to stay, and our purpose will be to classify rather than to abolish. Actually, my greater interest is in the development of a schema by which the methods may be classified, rather than in any particular inventory of procedures, but the schema proposed can be illustrated by tables of methods. And since the methods exist to solve the problems of psychophysics, it is appropriate to comment on what certain methods may and may not achieve in the way of solutions.

The Psychophysical Parameters

Psychophysics concerns the functional relation between stimulus and response: $R = f(S)$. This function is affected by numerous parameters. For the purpose of classifying methods it is convenient to distinguish three classes of these parameters, namely, the task undertaken by the observer, the manner in which the stimuli are presented, and the statistical measure employed in the description of the data. These classes and their principal subdivisions are listed in Table 1.

In Table 1 the various subdivisions of the three psychophysical parameters—task, stimuli, and statistic—are designated by a capital letter. These letters will be used in Table 2 to characterize the several psychophysical methods. But first let us examine the psychophysical parameters in a little more detail.

Task

The observer's task is normally set by means of instructions. The observer is "tuned" to react in one way rather than another, but from our present point of view only certain as-

TABLE 1

PSYCHOPHYSICAL PARAMETERS

Task of observer is to judge	Stimulus arrangement	Statistical measure
C Classification	F Fixed	L Measure of location (central tendency)
O Order	A Adjustable	V Measure of variability or confusion
I Intervals		
R Ratios		
M Magnitudes		

pects of this tuning are of consequence. It is important, for example, whether the observer is told to judge brightness, or hue, or saturation, but we are not here concerned with this aspect of the *Aufgabe*. Rather we take the attentional focus for granted and then ask what type of *relational* judgments the observer is trying to make. These relational judgments fall into five groups, as follows.

Classification (C). Here the observer's task is classification of one sort or another. He judges whether his perception meets some *nominal* criterion, with no reference to order among his perceptions. In the simplest case the observer, attending to some attribute or aspect of perception, judges whether it is *present* or *absent*. Thus he may press a key if he hears a tone, or if he hears a change in the tone. Or he may be required to say in what quadrant of a circle a light appeared, or in what interval of time a click was sounded. The task is that of detection, and the observer is set to behave as a yes-no device. In other cases the observer's task may be to judge *equivalence*, i.e., whether or not some criterion is met. The criterion may be set by a first stimulus, and the task may be to judge whether a second stimulus produces an equivalent effect, e.g., the loudness of one tone may be adjusted to match that of another tone. Or the criterion may be established by instruction, as when the subject is told to adjust a line to make it appear vertical, or to adjust wavelength to make a light appear pure green. Sometimes the classification problem is simply *identification* or *recognition:* is the present stimulus the same as some previous stimulus to which a name or number may have been assigned? Or, for example, was the sound presented an English word, and which word was it?

Order (O). The observer is set to judge greater or less, heavier or lighter, louder or softer, etc.

Intervals or distances (I). The observer's task is to judge apparent distance or difference between two or more perceptions. Ordinarily this takes the form of partitioning a continuum into apparently equal intervals or assigning stimuli to categories that seem equally spaced along a continuum.

Ratios (R). The observer attends to the relative magnitudes of two or more perceptions and reports the apparent ratios among them. Alternatively he may be set to produce stimuli that appear to stand in a prescribed ratio, which may be stated numerically or may be set in terms of some other pair of stimuli.

Magnitudes (M). The observer judges the apparent magnitude of a perception. He usually attempts to assign numbers proportional to the apparent magnitudes of a series of

stimuli. Or he may be set to produce stimuli that correspond to a series of prescribed magnitudes.

Stimulus Arrangement

Since there is no end to the variety of procedures that may be used for the presentation of stimuli, it might appear that no useful criterion for classifying them is possible. On the other hand, there is an important procedural distinction between confronting the observer with a variable stimulus and confronting him with a fixed stimulus. Variable or adjustable stimuli are better for some purposes, fixed stimuli for others. As a practical matter, fixed stimuli can nearly always be used, but adjustable stimuli, unfortunately, are not always easy to come by. It is hard, for example, to devise continuously adjustable weights for a lifted weight experiment, or continuously adjustable concentrations for experiments on taste.

We will distinguish then between two stimulus arrangements:

Fixed stimuli (F). Fixed stimuli are those that are not varied during the time they are being observed. Usually, of course, they are varied between observations.

Adjustable stimuli (A). Adjustable stimuli are those that may be altered during the course of observation. Usually the observer does the adjusting by operating a control, but the experimenter may operate the controls, or the adjustments may be made automatically, as in the method we shall call "tracking."

Statistical Measures

How the data from a psychophysical experiment are processed usually depends on the experimenter's purpose. Neglecting the secondary frills of statistical descriptions we can divide the usual treatments into two classes, depending on whether the final measure used is one or another measure of location (or central tendency, so-called) or one or another measure of variability, confusion, or dispersion.

Measures of location (L). For most purposes the measure we want to use is either a mean or a median. We want to know the typical response of a subject or of a group of subjects. Given a measure of location, a measure of dispersion can then be used to gauge the precision of the judgments.

The choice of a proper measure of location often presents interesting problems whose solutions are far from obvious. For one thing, we have a choice among such conventional measures as the mode, median, arithmetic mean, geometric mean, and harmonic mean. But it may turn out that none of these is appropriate. The arithmetic mean is inappropriate, for example, when applied to the readings obtained with a particular instrument whose indications are a nonlinear function of the quantity we want to average. Elsewhere (28) the writer has tried to suggest how an iterative procedure might aid in the solution of some of these problems. For a variety of reasons, a defensible rule concerning measures of location in psychophysical experiments is simply: when in doubt use the median. The median has the advantage that it is invariant under nonlinear transformations so long as they are increasing and monotonic.

Measures of variability (V). For our present purpose the measures of variability include the conventional measures of dispersion (standard deviation, average deviation, interquartile range, etc.), as well as measures of confusion, such as the proportions of times a given stimulus is judged in different categories. Measures of variability are often used in

the assessment of differential sensitivity or resolving power. They are also used, under various assumptions, as "distance" measures on psychological continua. This "unitizing" of dispersion may be legitimate on some types of continua, but on quantitative and intensitive continua the assumptions commonly made about discriminal dispersion are frequently in error (**31**). We will return to this problem later.

Problems and Methods

Some of the principal problems of psychophysics are listed in Table 2 along with some of the typical methods used in their solution. As is clear in Table 2, each of the problems of psychophysics may be regarded as one or another problem of scale construction construed in the widest sense of the term. This is scarcely surprising, for psychophysics, like most other parts of science, is mainly concerned with measurement. And since measurement is possible at different levels, ranging from nominal to ratio, the basic problems of psychophysics can be classified in a way that reflects these different levels.

In listing the problems in this way we must remind ourselves that they represent types or classes of issues that might concern the psychophysicist. Often the solution of one or another of these problems is not an end in itself, but is only a necessary step in the answering of a more far-reaching query. Especially in the more practical pursuits, such as human engineering, does it often turn out that the solving of problems like those in Table 2 is merely a means to an end. It has already been mentioned, for example, how the development of the sone scale of loudness has helped to answer a persistent problem facing the acoustical engineers (**29**, **33**). Even more extensive commercial applications are made of the Munsell color scales which are based on extensive psychophysical studies. Examples of this sort could be multiplied at length, but our interest here is more in fundamentals than in applications.

The methods listed under each problem do not exhaust the possibilities, nor do they all qualify necessarily as good procedures. On the other hand, these methods illustrate the procedures commonly used. The names given to the methods are mostly those in general use, although an occasional name is new, as is also an occasional method. In constructing tables of this sort one can scarcely avoid thinking of new procedures that ought to be put to test.

Let us now consider each problem in turn.

I. Nominal Scales

The nominal scale is the most general type of scale. It is the primitive variety that involves only classification, with no ordering or metricizing. Perhaps under some definitions this simple process may qualify, not as measurement, but as a kind of halfway house on the road to it. On the other hand, there is no doubt that the psychological processes involved in the forming of classes, concepts, or categories present rich and varied problems. As a matter of fact, Bruner, Goodnow, and Austin (**7**) have written a whole book on the problem of "categorizing and conceptualizing," which is essentially a problem in nominal scaling. So too are the manifold problems in such areas as pattern recognition and articulation testing.

In what follows, we will limit our interest to the three instances of nominal scaling that have long been central problems in psychophysics and to a fourth problem whose develop-

TABLE 2

PSYCHOPHYSICAL PROBLEMS AND METHODS

I. To determine nominal scales	
a. Absolute thresholds	
1. Single stimuli	CFL
2. Counting	CFL
3. Forced location (forced choice)	CFL
4. Adjustment	CAL
5. Limits	CAL
6. Tracking	CAL
7. Staircase (up-and-down)	CFL
b. Resolving power or differential sensitivity	
1. Adjustment (average error)	CAV
2. Tracking	CAV-OAV-CAL
3. Constant stimuli	OFV
4. Single stimuli	IFV-OFV
5. ABX	CFV
6. Forced location	CFL
7. Quantal increments	CFL
c. Equation of magnitudes	
1. Adjustment	CAL
2. Constant stimuli	CFL-OFL
3. Tracking	CAL-OAL
4. Staircase (up-and-down)	CFL-OFL
d. Identification	
1. Single stimuli	CFV
II. To determine ordinal scales	
1. Pair comparison	OFL
2. Rank order (order of merit)	OFL
3. Rating scale	OFL-IFL
4. Single stimuli	CFV
III. To determine interval scales	
1. Equisection (bisection)	IAL-IFL
2. Interval estimation	IFL
3. Category rating (equal intervals)	IFL
4. Category production	IAL
5. Pair comparison	OFV
6. Rank order	OFV
7. Successive categories	IFV
8. Successive intervals	IFV
IV. To determine logarithmic interval scales	
1. Pair comparison	OFV
2. Ratio matching	RAL-RFL
V. To determine ratio scales	
1. Ratio estimation	RFL
2. Ratio production (fractionation, multiplication)	RAL-RFL
3. Magnitude estimation	MFL
4. Magnitude production	MAL

NOTE.—The capital letters after each method refer to the psychophysical parameters in Table 1. Alternative procedures under a given method are indicated by multiple sets of letters.

ment stems from information theory. Although conventional treatments do not usually subsume these problems under the heading of nominal scales, my colleague, Ulric Neisser, has pointed out to me that that is where they belong. Each of them reduces in one way or another to the classification of stimuli. Thus, in measuring absolute thresholds we form a twofold classification: those stimuli that can be perceived and

those that cannot. Similarly, in measuring resolving power or differential sensitivity we divide stimulus increments into two classes, detectable and not detectable. In the equation of magnitudes our task is obviously to form the class of equivalent stimuli (according to a particular criterion) and to set if off from the class of stimuli that are not equivalent. And in the fourth problem our interest may be to determine into how many distinct classes a person can divide a set of stimuli without confusing any of them. Or it may concern the recognizability of stimulus configurations and the parameters that govern such recognition. In trying to solve these problems, we confront the observer with the task of putting stimuli into classes. The psychophysical procedure involved requires only classification and does not require that the observer order his perceptions or judge the intervals or ratios among them.

It may be true, of course, that under a particular experimental procedure the observer may be asked to judge "greater or less," but if the problem is really one of nominal scaling (e.g., the measurement of resolving power) the experimenter will proceed to use the experimental results to determine class boundaries. In other words, the categories of the nominal scale will be abstracted from the observer's judgments of apparent order.

Absolute Thresholds

The absolute threshold, the point that divides the continuum of stimuli into those the subject can detect and those he cannot, tends to elude our efforts to define its locus because it shifts about in time and we are forced to trap it by sampling and statistics. Mostly we use one or another version of the method of *single stimuli*. Fixed stimuli are presented at various levels and the yes-no responses of the subject are recorded. The class boundary which we call the threshold may be defined as the stimulus level detected half the time.

The method herein called *counting* is a procedure sometimes used in the large-scale testing of inexperienced subjects. Several stimulus levels are presented in a series and the subject reports how many he perceived. Other variations on this method may call for the presentation of two or more stimuli at a fixed level.

The method sometimes called *forced choice* (**5**, **40**) is not unlike counting except that, instead of asking the subject to say how many stimuli were presented, we ask him to say where in space or in which of several intervals of time the stimulus occurred. Since many methods call for a forced choice (e.g., constant stimuli used with two-category response), a better name for this procedure might be *forced location*. The subject tries to locate the stimulus in space or time. Blackwell has studied many of the parameters of this process.

The foregoing methods all make use of fixed stimuli. An adjustable stimulus can be used under the method of *adjustment*, in which the observer sets the level to be "barely detectable." This procedure is quick and convenient, but ordinarily it allows only a rough determination of the threshold.

If the level is varied systematically from points below and above threshold, and the subject signals when the threshold is crossed, we call the procedure the method of *limits*.

The term *tracking* is suggested to designate a procedure made popular by the Békésy audiometer (**2**, **3**). The observer presses a key whenever he hears a tone. As long as the key is pressed the level of the tone de-

creases, and when the key is released the level rises. By this means the subject may track his threshold throughout the frequency range, and by means of a recording pen the track of the stimulus is traced across an appropriate grid. A smoothed curve through the zigzag track is usually drawn to depict the threshold locus.

A method of tracking was apparently developed independently by Oldfield who used it for the measurement of visual thresholds (**18**). The method of tracking has also been used with animals. Blough (**6**) devised an experiment in which a pigeon was trained to peck at one key when a target was visible and at another key when the target was too dim to be seen. The pecks were made to control the brightness of the target, and the pigeon was thereby able to track its own dark-adaptation curve.

The *staircase* method, sometimes called the "up-and-down" method (**8**), is much like the method of tracking except that the stimulus level is not varied continuously. The levels used are fixed and discrete, and the level presented on a given trial depends on the response made on the previous trial. Thus if the previous stimulus was detected, the level of the next stimulus is lowered a step; if the previous stimulus was not detected, the level is raised.

Resolving Power

The measurement of the difference limen, ΔS, has been a major concern of classical psychophysics, and enormous labors have gone into the refinement of methods. Yet there is no agreed-upon best procedure for determining the least resolvable difference between two stimuli. Differential sensitivity is a difficult, "noisy" thing to measure, and it is not surprising that different procedures give different results.

Measures related to resolving power are sometimes obtained as a kind of by-product from such methods as *adjustment* and *tracking*. Some measure of dispersion about the mean adjustment or the "center" of the track may be used to define the just noticeable difference. In principle, measures of resolving power may be derived from the scatter of observations obtained from a wide variety of psychophysical procedures.

The method of *tracking* has also been adapted to the direct measurement of just noticeable differences (**19, 42**). The trick here is to modulate the level of a signal alternately up and down, and to let the observer's responses control the amount of the modulation. He presses a key whenever he detects modulation and releases it when the level of the stimulus no longer appears to rise and fall. When the key is pressed the degree of modulation declines and when the key is released the degree of modulation increases. Thus the observer is able to "track" the just noticeable *change* in the signal. The average detectable modulation is determined by a curve drawn through the center of the zigzag track. This procedure has been used successfully with auditory stimuli, and there appears to be no reason why it should not lend itself equally well to some other types of stimuli.

Although it is not listed in Table 2, the staircase adaptation of the method of tracking could also be used to determine the threshold of a modulation.

Note that in Table 2 the method of tracking is scored CAV-OAV-CAL. This is intended to suggest that the observer may be instructed to judge in different ways, or that different statistical measures may be used to define the differential threshold. Other double scorings in Table 2

stand for alternative procedures under a given method.

The three methods, *constant stimuli*, *single stimuli* and *ABX*, all have much in common. Constant stimuli employs a standard and a series of comparison stimuli, and the observer judges ordinarily in terms of greater or less. Single stimuli (sometimes called absolute judgment) dispenses with the standard and the subject judges in terms of two or more categories, such as light, medium, or heavy. In the ABX method the subject reports (forced choice) whether the third stimulus is more like the first or the second of a series. In one study this method appeared to yield larger difference limens than the method of constant stimuli with a two-category forced-choice judgment (**21**).

Under the method of *forced location* increments are added to a steady stimulus and the subject is required to say when in time, or where in space, the increment occurred (**5**). The measure of sensitivity may be taken as the increment correctly identified half the time, after correction for chance.

The method of *quantal increments*, sometimes called the *quantal procedure* (**38**), attempts to measure the size of the stimulus increment needed to produce an all-or-none jump in effective excitation. As developed by Stevens and Volkmann (**39**), the procedure calls for a steady stimulus to which brief increments are added periodically. The observer presses a key whenever he perceives an increment. Under optimal conditions we obtain a rectilinear psychometric function of predictable slope, from which the size of the "neural quantum" can be gauged in terms of stimulus units. Whenever the possibilities for stimulus control are such that this method can be used, the writer thinks it is the preferred procedure. Psychometric functions of the predicted form have been obtained for pitch and loudness, and the data Mueller (**17**) obtained for brightness fit the linear functions predicted by the quantal hypothesis as well as, if not better than, they fit the sigmoid curves Mueller drew through them. And the slopes of some 45 psychometric functions all cluster about the predicted slope.

On the other hand, some experimenters have apparently not succeeded in reducing the variability and "noise," whether in the experimental procedure or in the observers, to the point of obtaining clear "quantal functions." The task of proving the stepwise character of discrimination is not easy. It is in some ways like trying to prove that the charge on the electron is constant. Prior to Millikan's oil-drop experiment, several attempts seemed to show that the charge was not constant but was probably normally distributed. For the electron the physicist now knows pretty well how to set up a repeatable experiment that will demonstrate the all-or-none nature of the charge, but as yet, unfortunately, we cannot prescribe all the conditions that will guarantee a "quantal" psychometric function. One reason is that the observer is too important a part of the specification. The writer can name several friends and colleagues who have been able to hold the steady attention needed for this task, but he does not know how to specify the differences between them and the observers who have not done so well at it.

Nevertheless, whether or not good clear quantal functions are obtained in any given experiment, the procedure itself has much to recommend it for the purpose of mapping resolving power. The method of quantal

increments goes directly at the problem of what increments can be detected, and it provides for internal checks on the "noisiness" of the results obtained.

Equation of Magnitudes

The determination of equivalence is the problem we try to solve when we map such things as equal loudness contours, luminosity functions, contours of constant hue, etc. We try to determine the class of stimuli that appear equal with respect to a particular attribute. It resembles what the economist does when he maps indifference curves for utility (**35**).

For the typical matching problem the method of *adjustment* is usually the speediest and most straightforward. It is not, however, without its constant errors (**29**). The method of *constant stimuli* is also widely used for this purpose, and it too has its constant errors, in particular, the so-called "time-error" (**31**).

The method of *tracking* can also be used to trace out an equivalence contour. Zwicker and Feldtkeller (**41**) presented two tones alternately. One was of fixed intensity and frequency (1000 cps), and the other was made to sweep slowly through the frequency range. The observer pressed a key whenever the variable tone sounded louder than the standard and released the key whenever the variable sounded fainter. While the key was pressed, a motor-driven attenuator decreased the level of the variable tone, and when the key was released the motor reversed its direction and the variable grew louder. The data appear as a zigzag line traced by a pen across an audiometric chart, and an "average" line drawn through the zigzag tracing shows directly the form of the equal loudness contour.

The staircase or "up-and-down" method is an adaptation of the method of tracking for use with fixed values of the comparison stimulus. Applied to the problem of equating magnitudes, the staircase method is a kind of cross between constant stimuli and tracking. It is like constant stimuli in that a standard and a set of fixed comparison stimuli are used, but it is like tracking in that the response of the subject determines the value of the subsequent comparison stimulus. For example, if the subject says "greater" the next comparison stimulus is decreased; if he says "less" the next comparison stimulus is increased.

The mapping of invariances by means of a matching procedure is the basis of much of what we sometimes call measurement in psychophysics. When no other scales are available we often "measure" the effect of one factor in perception by changing that factor and then finding what alteration of another parameter will appear to undo the change. Thus we measure the effect of intensity on pitch by changing the intensity a given amount and then altering the frequency to restore the pitch to its original state. Or the observer may adjust the intensity instead of the frequency (**24**). We then measure the effect of intensity in terms of the frequency change required to cancel the effect of the change in intensity. Examples of this sort could be multiplied at length.

Identification

In the course of its burgeoning development, "communication theory" has had numerous impacts on psychology. Despite the fact that the theory deals only with the nominal properties of ensembles, not with their ordinal or interval properties, measures of information have found a use in many types of inquiry. And

since the theory provides a mathematical model that concerns association at the nominal level of scales, it is not surprising that it should be put to work in psychophysics.

A typical problem in this area concerns the ability of observers to identify or recognize a stimulus in an absolute sense, without confusing it with other stimuli. The problem is related, of course, to differential sensitivity, but the concern is more with correct naming than it is with resolving differences. The distinction is something like that between pitch discrimination on the one hand and so-called absolute pitch on the other. Absolute pitch concerns the ability of a listener to name or identify the note played to him.

For many purposes it is important to know how many different stimuli (e.g., frequencies, intensities, colors, pressures, etc.) a person can identify with minimal error and to determine what factors affect information transmission on the different sensory continua (**1**). It turns out that, on many of the common sensory continua, perfect transmission of information (with no confusions) is ordinarily not possible with more than about five different stimuli (**16**). This fact is basic to many practical problems in the coding of information.

Also related to such problems is the question concerning the best way to distribute stimuli along a continuum in order to maximize information transmission. Studies of this sort have led to what is sometimes called an "equal discriminability scale" (**10**). Since the subject's task in these experiments is to identify and not to make comparisons, perhaps a better name would be "equal identifiability scale." In any case the psychophysical method employed in these problems is *single stimuli*. The experimenter presents a stimulus and the subject tries to give the appropriate response.

Problems of this sort are of course not limited to single stimulus dimensions. The problem of pattern recognition, for example, nearly always involves multivariate stimuli, usually visual or auditory although they may be tactual (**12**), and our interest is in how the observer is able to classify or recognize complex stimulus configurations. As already noted, it is to these problems that information theory has contributed useful tools.

II. Ordinal Scales

For the setting of perceptions in a rank order with respect to some aspect or attribute we have three conventional methods: *pair comparison*, *rank order*, and *rating scale*. The use of these methods for the purpose of ordering is usually straightforward and devoid of special problems. In the sensory area the ordering of psychological magnitudes is seldom a serious problem, because subjective magnitude is usually a monotonic function of stimulus magnitude, and we take the ordering for granted. Occasionally, however, we find that the simple monotonic relation breaks down. For example, the apparent saturation of a light of a single wavelength grows with intensity up to a certain value, but when the intensity is increased further the color blanches out and saturation declines.

The problem of ordering is sometimes extended to the relative *spacing* among stimuli, and the order of the spacing is sometimes deduced from confusions among the stimuli. The reasonable assumption is made that, if stimulus B is confused with A more often than C is confused with A, then C is farther from A than B is from A. The method of *single stimuli* is one procedure that might be used in an experiment of this sort.

Although measurements of confusion may order stimuli relative to a single point, as in the foregoing example, there are situations in which relative distances cannot be ordered by measures of confusion. On prothetic continua, where jnd's do not constitute subjectively equal distances (**31**) it is not always true that distances can be ordered by this procedure. Thus, if A and B are confused more often than C and D, it does not always follow that the distance from A to B is less than from C to D. In particular, if A and B are two intense tones, and C and D are two faint tones (all of the same frequency), tones A and B may be farther apart in subjective magnitude (sones) and still be confused more often. For example, tones of 100 and 101 sones would be confused more often than tones of 1.0 and 1.5 sones.

III. Interval Scales

Many interesting problems arise when we take on the task of erecting a scale of equal intervals on a psychological continuum. The determination of equal sense-distances seems to have originated with Plateau, who asked eight artists to paint a gray whose shade appeared equidistant between a black and a white (see **31**). In this manner, Plateau invented the method of *bisection*, which became *equisection* when the intervals were subdivided further. But Plateau, like many after him (see, for example, **13**), did not perceive the basic difference between bisection and such "ratio methods" as fractionation. As we shall see, this distinction is very important for two reasons: (*a*) bisection can lead at best only to an interval scale, and (*b*) it turns out that human observers are so constituted that they are generally unable to bisect an interval on a quantitative or intensitive (prothetic) continuum without making a systematic error.

The fact that bisection leads only to an interval scale is obvious enough. That subjects cannot perform valid bisections on certain types of continua is not so obvious, however. Evidence for this statement is described elsewhere (**31**), but since the argument is relevant to our present concern, let us review it briefly.

Systematic studies of more than a dozen perceptual continua have shown that these continua divide themselves into two varieties. Class I comprises the quantitative, intensitive continua, the continua concerned with *how much*. Discrimination on certain of these continua, such as loudness, brightness, and heaviness, seems to involve an additive process at the physiological level. For this reason Class I has been called *prothetic*. Class II includes the qualitative and positional continua, the continua concerned with *what* or *where*. Discrimination on these continua seems to involve a substitutive process, and they are therefore called *metathetic*.

Now, on prothetic continua we find that the psychological magnitude, as determined by ratio scaling procedures, approximates a power function of the stimulus magnitude. The rule is that equal stimulus ratios correspond to equal sensation ratios. This rule, the writer has suggested, is a basic "psychophysical law." The exponents of the power functions range from about 0.3 for loudness and brightness to about 3.5 for the apparent intensity of electric current applied to the fingers (**34, 36**). Since differential sensitivity on these continua tends to approximate Weber's law, or rather the modified form of this law, $\Delta S = k(S+c)$, it follows that the psychological magnitude represented by ΔS increases as the stimulus

increases. Discrimination, in other words, is not constant over the continuum, when measured in subjective units. There is a basic asymmetry in sensitivity.

On the other hand, on metathetic continua, discrimination, measured in subjective units, tends to be uniform over the scale, and there is no systematic asymmetry as there is with prothetic continua. This difference between the two continua makes for rather different behavior on the part of the observer. On the "symmetrical" metathetic continua, the results of bisection tend to agree with the results obtained by direct magnitude estimation and related procedures. But on the "asymmetrical" prothetic continua the point of bisection tends systematically to be lower than the point predicted by direct magnitude estimation. (Bisection is also plagued by a curious and dramatic order effect, which has been called "hysteresis" [**31**].)

The phenomena that characterize equisection also show up in the method of *interval estimation*, which is a kind of inverse of equisection. In equisection the intervals are adjusted to meet some criterion (usually equality), whereas in interval estimation the experimenter sets a series of stimuli and the observer estimates their apparent spacing. A convenient procedure for reporting these estimates is to have the observer adjust the positions of a set of markers along a line. The apparent intervals between successive markers are made to appear proportional to the apparent intervals between the stimuli. This method has been used with loudness and with lifted weights (**31**). The markers were movable sliders on a steel bar set before the observer. This procedure is in some ways analogous to the use of a continuous rating scale on which the judge places a pencil mark on a line. Since apparent position (on a line) is a metathetic continuum on which discrimination is not asymmetrical, adjustments of visual position provide an unbiased method of assessing the apparent spacing of other stimuli—except for possible distortions due to end effects (**37**). Thus the essential features that characterize equisection on prothetic continua (hysteresis and the bias due to the asymmetry of sensitivity) are also revealed by the method of interval estimation. (For an interesting variation on the method of interval estimation, see **17***a*).

The discrepancy between the "interval" judgment and the judgment of magnitude is especially striking when we use the method of *category rating*, under which the observer assigns a finite set of numbers or adjectives to a set of stimuli and tries to space the categories equally. Plotted against the ratio scale of subjective magnitude, the category scale has turned out to be concave downward on nine prothetic continua recently examined (**36, 37**). This nonlinearity in the category scales shows up even when the "pure" form of the category scale is obtained by a process of experimental iteration.

On metathetic continua, such as pitch, position, inclination, and proportion, the category scale *may be* linearly related to the magnitude scale, provided the distortions due to stimulus spacing, landmarks, and differential familiarity have been neutralized. These factors are some of the second-order variables that can alter the form of the category scale.

Another method for obtaining a category scale is the method of *category production* (**37**). This is a kind of inverse of category rating. Instead of asking the observer to assign categories to the stimuli, the experimenter names the categories, in irreg-

ular order, and the observer adjusts the stimulus to produce his conception of each category. Examples of the extreme categories (e.g., No. 1 and 7) may be presented to the observer at the outset. In our few tests of this method we found that it seemed to give directly a close approximation to the "pure" category scale that would be obtained by experimental iteration.

What is essentially the method of category production has also been used to study how people make linear interpolations in a spatial interval (**22**). This is a metathetic continuum. The experimenter is usually concerned with the "objective" accuracy of the observer's settings, although he may also be concerned with the form of the observer's subjective scale.

In summary, then, the four methods, equisection, interval estimation, category rating, and category production, are all designed to produce an interval scale of "equal sense-distances." If properly used, they can achieve this end on metathetic continua, but on prothetic continua they fail to produce intervals that are equal, as measured by the ratio scales of the continua.

We turn now to the class of methods that seek to produce equal intervals via the "unitizing" of one or another measure of variability. *Pair comparison* is perhaps the best known example of this procedure, but the underlying philosophy is similar for the other three methods listed in Table 2 (see **14**). By making certain simple assumptions regarding the distribution of the observed variabilities or confusions, we try to deduce the form of the underlying continuum. It would appear that on metathetic continua, where sensitivity to differences is uniform (in subjective units), the distribution assumptions most commonly invoked may lead to an interval scale. But on prothetic continua, the assumptions of normal and uniform variability are demonstrably in error, and therefore the resulting scales are not scales of equal intervals. On prothetic continua the procedures ordinarily used to derive equal intervals from measures of variability or confusion miss the mark for the same reason that "Fechner's law" fails: the subjective size of the jnd is not constant over the continuum. Likewise "discriminal dispersion" is not uniform over a prothetic continuum. As a matter of fact, the psychophysical power law, coupled with the relativity of resolving power, leads us to predict that discriminal dispersion is not distributed normally (or even symmetrically) on a linear subjective measure of a prothetic continuum, although it may be normal on a logarithmic measure of the continuum.

We must conclude, I think, that those procedures that make use of an assumed canonical distribution of variability are less useful for scaling than methods that utilize directly a measure of location. Even so, in the determination of equal intervals on prothetic continua, these latter methods are themselves subject to invalidating biases. It appears, therefore, that the only proper method for determining equal intervals on a prothetic continuum is to construct a ratio scale (see Section V). This solution is possible because the ratio scale contains the interval scale.

IV. Logarithmic Interval Scales

The possibility that discriminal dispersion may increase proportional to the psychological magnitude on a prothetic continuum suggests that an assumption to this effect might make it possible to scale the continuum into intervals that are equal in terms

of logarithms. So far as the writer is aware, no use has been made of such scales, although he has elsewhere (**31, 35**) described some of their properties, including their mathematical group structure. In proceeding in this fashion we would be assuming that the conventional procedures used to scale a continuum by the method of pair comparison give us, for prothetic continua, a scale on which the values are separated not by equal intervals, but by equal ratios, i.e. $a/b = b/c = c/d = \ldots$.

The equating of ratios, either by way of a processing of variability or by a direct judgment of apparent ratios, would provide the basis for what the writer has called a *logarithmic interval* scale. This scale is invariant under a power transformation, i.e., for any value x we can substitute x' where $x' = ax^b$, and where a and b are positive numbers. As with the linear scale of equal intervals, the zero point on the logarithmic scale can be chosen arbitrarily and moved at will.

If such a scale were desired, the straightforward procedure for achieving it would presumably be some procedure of direct *ratio matching*. Methods of this type have not been very thoroughly explored, although Garner (**11**) tried equating loudness ratios. He seemed to find evidence that observers may not be able to keep separate the two tasks, that of equalizing ratios and that of equalizing intervals. On the other hand, J. C. Stevens obtained interesting results when two brightnesses were used to define a subjective ratio, and the observer adjusted the ratio between two loudnesses to make the ratio between the loudnesses match the apparent ratio between the brightnesses (**31**).

Ratio matching of this kind may have utility for psychophysics, because it provides an alternative method for demonstrating that the "psychophysical law" governing prothetic continua is a power function, i.e., the psychological magnitude is equal to the stimulus magnitude raised to a power. In principle, the power function can be tested by this method without requiring the observer to make numerical estimates of ratios or of magnitudes (as described in the next section). The observer is required only to make the apparent ratio between one pair of stimuli equal the apparent ratio between another pair of stimuli. Then if it turns out that the ratio of one pair always equals the ratio of the other pair raised to a power, it follows that the psychological magnitudes are power functions of their respective stimuli.

For example, suppose the subject adjusts pairs of luminances (B_1 and B_2) to make the ratio of the brightnesses equal the ratio of the apparent lengths of two lines (L_1 and L_2). If for all possible pairs so matched we find

$$L_1/L_2 = (B_1/B_2)^n$$

where n is a constant, then both subjective length and subjective brightness are power functions of their respective stimuli. The limitation on this approach is that we could not by this procedure determine the value of the two exponents involved. That would require the additional methods described below. We could, however, by such ratio matchings determine the *relative* values of the exponents for length and brightness.

Whether observers can make such ratio matches with sufficient consistency to make it a profitable procedure has not been explored very thoroughly. Nevertheless, the experiments involving the matching of ratios of loudness and brightness

have thus far been encouraging. In this particular case, the observers adjusted the two sounds to approximately the same physical ratio as the experimenter set between the two luminances. This outcome is consistent with other evidence that loudness and brightness are both power functions of their stimuli, not logarithmic functions as Fechner supposed, and that the exponents for both loudness and brightness are approximately the same. By means of the ratio scaling procedures discussed below we have shown that both exponents are of the order of 0.3 (**31**).

V. Ratio Scales

Perhaps the lack of interest in logarithmic interval scales stems from the fact that the scientist's greater interest lies in ratio scales. He is less interested in ratios whose values are equal but indeterminate than he is in ratios whose values he can specify. If he has a procedure for equating ratios, plus a procedure for equating intervals, he can proceed to construct a ratio scale on which the zero point is not arbitrary (**26**).

We have seen, however, that observers exhibit a systematic bias when they try to equate intervals on a prothetic continuum. Whether, and under what conditions, observers can equate a series of unknown ratios has not been fully explored. In view of this state of affairs, how do we create ratio scales on a perceptual continuum?

The answer seems to be that we ask the subject to judge the value of the ratio, or of the magnitude, directly. Using one or another, or preferably combinations, of four different methods we proceed directly to the goal of assessing relative psychological magnitudes. The potential methods are as follows:

Ratio estimation calls for the presentation of two or more stimuli and the observer names the value of the apparent ratio between them. The so-called *constant sum* method (**15**) is a special instance of this procedure. As typically used, this method requires that the observer divide 100 points between two stimuli in such a way that the division between the points reflects the apparent ratio between the sensations. In general, however, there is little reason to restrict the observer's method of report in this manner. Restrictions and constraints on the observer are often a source of trouble and bias in ratio and magnitude judgments.

It should be mentioned that, in using any of the methods for determining ratio scales, the experimenter will generally do well to compute medians, for there is no limit to how far an occasional observer may deviate from the rest of the group.

Ratio production is probably best known by the name of one of its subvarieties, *fractionation*. The observer adjusts a stimulus to produce a prescribed ratio between two apparent magnitudes. He sets a variable to be $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, or some other fraction of a standard. Alternatively the experimenter may set the stimuli and the observer may report whether they meet the criterion of a prescribed ratio. This procedure is analogous to the method of constant stimuli and is sometimes called by this name. The principal drawback of fractionation with fixed stimuli is that the choice of the levels at which the comparison stimuli are set may be critical. This difficulty can presumably be surmounted if the spacing of the comparison stimuli is determined by a process of experimental iteration under which the levels are altered in accordance with the outcome of successive experiments in such a way

that the "criterion" stimulus is made to lie at or near the center of the range of the comparison stimuli (**37**).

The method of *fractionation* has its inverse in the method of *multiplication* under which the observer sets a variable to some prescribed multiple of a standard. In order to balance out certain sources of bias it is nearly always wise to complement fractionation with multiplication (**32**). This is especially true when wide ranges of stimulus values are explored in experiments on loudness and brightness.

Magnitude estimation refers to a procedure by which the observer makes a direct numerical estimation of the psychological magnitudes of a series of perceptions (**30**). Two main varieties of this procedure have been used. Under one of them the experimenter presents a stimulus and assigns it a number (modulus) such as 10, say. He then presents other stimuli, and the observer assigns to them numbers proportional to their apparent magnitude. Under the other variation in the method, no modulus is prescribed. The stimuli appear in irregular order and the observer assigns numbers proportional to magnitude, using a modulus of his own choosing.

The method of direct magnitude estimation has given good results with loudness, brightness, lifted weights, duration, lightness of grays, visual length, pitch, proportion (**37**), finger span (unpublished), vibration (unpublished), and electric shock (**34, 36**). Incidentally, with this method it often turns out that the geometric mean falls close to the median.

Magnitude production is a method that has been named but not yet thoroughly explored (**31**). It is the inverse of magnitude estimation in that the experimenter states a series of magnitudes (presumably in irregular order) and the observer adjusts a stimulus to produce them. In some ways the procedure resembles the method of category production described above, except that no range is specified and the observer tries to judge in terms of apparent magnitudes and not in terms of a finite number of prescribed categories. Magnitude production is a potentially interesting method, provided the stimulus control is such that the subject can adjust the stimulus over the required range, but many questions concerning its peculiarities and difficulties remain to be answered. It is not unlikely that the biases in magnitude production are such that, in a balanced program, they might be used to offset some of the systematic errors in magnitude estimation. It seems, in general, that each of the ratio-scaling methods may contain biases peculiar to itself, and that the elimination of the biases can sometimes be achieved by means of a counterbalanced design in which the biases inherent in one method are evaluated and corrected by means of a method that contains biases of an opposite sort. The principle is analogous to that employed in the use of a balance in weighing an object: in order to discover and correct for possible asymmetries in the balance we interchange the weights on the scale pans.

We have already noted that, on prothetic continua, ratio scales of psychological magnitude turn out to be power functions of the stimulus magnitude. The power law seems to hold on at least 16 perceptual continua. On metathetic continua, the ratio scaling methods may or may not give a power function. When a power function is found to hold, the exponent appears to be 1.0. But on a metathetic continuum like pitch, the psychological magnitude, measured

in mels, is definitely not a power function of frequency in cycles per second (**37**).

Another important point to note is that, when prothetic continua are scaled by the ratio methods, Fechner's law obviously fails. The relation between perceptual magnitude and stimulus magnitude is not logarithmic. Even if we take the weaker form of Fechner's law, which says that the counting off of jnd's gives the scale of perceptual magnitude, we find that the law fails almost as badly. Successive jnd's, in other words, are not subjectively equal.

On metathetic continua, on the other hand, the jnd scale apparently coincides with the scale of psychological magnitude. Thus, to a fair approximation, jnd's in pitch represent constant increments in mels (**27**).

The Estimation of Objective Values

The methods of psychophysics are ordinarily designed to solve problems related to the nature of organisms. The focus of interest is typically the normal observer, his thresholds, his resolving powers, and the magnitudes of his perceptions. The methods have, of course, their clinical uses, and the assessment of individual differences is central to many practical undertakings.

Quite different in aim, but sometimes similar in procedure, is another human activity involving discrimination and judgment. This is the use of the human being as an instrument to measure the objective values of things. Despite the ingenuity of modern instrumentation, many tasks of rating, grading and judging can still best be done by two-legged meters (cf. **23**). Man's sensing, differentiating, and integrating circuits still surpass in flexibility and power any inanimate substitutes yet devised. Instruments may aid but they do not displace the wine taster, the leather grader, the lumber sorter, or any of a host of other judges on whom commerce depends for the appraisal of its wares. Little of this type of activity gets attention in the academic laboratory, although much could probably be learned from its systematic study.

In the framework of our present concerns, the assessing and grading of objective things are practical problems—a potential field perhaps for an applied psychophysics. The chief difference between problems of this type and those we have been considering lies in the point of view. In addressing any of the five problems in Table 2 we seek to learn the properties of the human instrument; in problems of grading we care nothing about the properties of the instrument as such, but only about the accuracy of its indications. In the grading of wool, for example, the mill owner hopes the assessment of the "clip" will tell him more about the wool than it does about the grader. He hopes in other words, that the grader will commit the "stimulus error" 100%. The psychophysicist presenting a series of tones to be judged for loudness hopes quite the opposite. He wants the subject to report apparent loudness and not to judge how many decibels are probably being produced by the earphone. Some experienced judges can do either task at will, but the estimation of decibel levels requires a lot of training. How the properties of the judge, as appraised and systematized by psychophysics, interact with the applied problems of grading and rating is a potentially interesting problem.

The capacity of the human instru-

ment to make correct assessments of this or that is also a central issue in some branches of engineering psychology, and it is in these connections that the most systematic studies have been made. This is not the place to review this lively field, but an illustrative example may be in order.

It concerns the continuous control of a complex system, like the problem faced by the pilot taking a large ship through the narrow channels of the Suez Canal. The same kind of problem comes home to many of us when we try to back an automobile with a luggage trailer tied on behind. We watch the trailer to see where it is going. It goes to the left, let us say, when it should go to the right, so we turn the front wheels of the car and watch to see what happens. The trailer corrects its course to the desired direction, whereupon we straighten out the front wheels and continue to back up. To our dismay the trailer keeps turning to the right. We find we have overcorrected, because we attended only to the "output" of the system, and did not take into account the delay between the input control (turning the front wheels) and the final movement of the trailer. With enough practice, a driver can learn to manage the procedure and even to back the trailer into a garage. He must learn not only to judge the position and direction of the trailer, but also to act as an integrator and predict the effects of his control actions as they will be summed up over a period of time.

Problems of this sort have opened new fields of study concerned with man-machine control systems (**4**). One of the primary problems is to learn what objective aspects of the situation the operator can judge with greatest reliability, and then to display to him only those features that match his judgmental capacity. It has been found, for example, that best results are achieved when the human operator is relieved of the task of performing integrations and differentiations. The controls and displays must be engineered in such a way that the operator can effectively control the system by acting as a simple amplifier. The need for a complex and difficult judgment of objective values is thereby replaced by a simpler demand on judgment.

REFERENCES

1. Alluisi, E. A. Conditions affecting the amount of information in absolute judgments. *Psychol. Rev.*, 1957, **64**, 97–103.
2. von Békésy, G. Über ein neues Audiometer. *Arch. elekt. Übertragung*, 1947, **1**, 13.
3. von Békésy, G. A new audiometer. *Acta. Oto-Laryng. Stockh.*, 1947, **35**, 411–422.
4. Birmingham, H. P., & Taylor, F. V. A design philosophy for man-machine control systems. *Proc. I.R.E.*, 1954, **42**, 1748–1758.
5. Blackwell, H. R. Psychophysical thresholds. *Engng Res. Bull.* No. 36, Engineering Research Institute, Univer. of Michigan, 1953.
6. Blough, D. S. Method for tracing dark adaptation in the pigeon. *Science*, 1955, **121**, 703–704.
7. Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.
8. Dixon, W. J., & Massey, F. J., Jr. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
9. Evans, R. M. *An introduction to color.* New York: Wiley, 1948.
10. Garner, W. R. An equal discriminability scale for loudness judgments. *J. exp. Psychol.*, 1952, **43**, 232–238.
11. Garner, W. R. A technique and a scale for loudness measurement. *J. acoust. Soc. Amer.*, 1954, **26**, 73–88.
12. Geldard, F. A. Adventures in tactile literacy. *Amer. Psychologist*, 1957, **12**, 115–124.

13. GRAHAM, C. H. Behavior, perception and the psychophysical methods. *Psychol. Rev.*, 1950, **57**, 108–120.
14. GUILFORD, J. P. *Psychometric methods* (2nd ed.) New York: McGraw-Hill, 1954.
15. METFESSEL, M. F. A proposal for quantitative reporting of comparative judgments. *J. Psychol.*, 1947, **24**, 229–235.
16. MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.
17. MUELLER, C. G. Frequency of seeing functions for intensity discrimination at various levels of adapting intensity. *J. gen. Physiol.*, 1951, **34**, 463–474.
17*a*. NEWHALL, S. M. A method of evaluating the spacing of visual scales. *Amer. J. Psychol.*, 1950, **63**, 221–228.
18. OLDFIELD, R. C. Continuous recording of sensory thresholds and other psychophysical variables. *Nature*, 1949, **164**, 581.
19. OLDFIELD, R. C. Apparent fluctuations of a sensory threshold. *Quart. J. exp. Psychol.*, 1955, **7**, 101–115.
20. OPTICAL SOCIETY OF AMERICA, COMMITTEE ON COLORIMETRY. *The science of color.* New York: Crowell, 1953.
21. ROSENBLITH, W. A., & STEVENS, K. N. On the DL for frequency. *J. acoust. Soc. Amer.*, 1953, **25**, 980–985.
22. SCHUBERT, R. G., & JENKINS, W. L. The effect of brief training on linear interpolation. *J. appl. Psychol.*, 1956, **40**, 53–54.
23. SCOTT BLAIR, G. W. *Measurements of mind and matter.* New York: Philosophical Library, 1956.
24. STEVENS, S. S. The relation of pitch to intensity. *J. acoust. Soc. Amer.*, 1935, **6**, 150–154.
25. STEVENS, S. S. Sensation and psychological measurement. In E. G. Boring, H. S. Langfeld, and H. P. Weld (Eds.), *Foundations of psychology.* New York: Wiley, 1948.
26. STEVENS, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
27. STEVENS, S. S. Pitch discrimination, mels, and Kock's contention. *J. acoust. Soc. Amer.*, 1954, **26**, 1075–1077.
28. STEVENS, S. S. On the averaging of data. *Science*, 1955, **121**, 113–116.
29. STEVENS, S. S. The calculation of the loudness of complex noise. *J. acoust. Soc. Amer.*, 1956, **28**, 807–832.
30. STEVENS, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.
31. STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.
32. STEVENS, S. S. Concerning the form of the loudness function. *J. acoust. Soc. Amer.*, 1957, **29**, 603–606.
33. STEVENS, S. S. Calculating loudness. *Noise Control*, 1957, **3**, No. 5, 11–22.
34. STEVENS, S. S. Measurement and man. *Science*, 1958, **127**, 383–389.
35. STEVENS, S. S. Measurement, psychophysics, and utility. In *Symposium on measurement: held by the AAAS, December 1956.* New York: Wiley, in press.
36. STEVENS, S. S., CARTON, A. S., & SHICKMAN, G. M. A scale of apparent intensity of electric shock. *J. exp. Psychol.*, in press.
37. STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, **54**, 377–411.
38. STEVENS, S. S., MORGAN, C. T., & VOLKMANN, J. Theory of the neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.*, 1941, **54**, 315–355.
39. STEVENS, S. S., & VOLKMANN, J. The quantum of sensory discrimination. *Science*, 1940, **92**, 583–585.
40. TANNER, W. P., & SWETS, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.
41. ZWICKER, E., & FELDTKELLER, R. Über die Lautstärke von gleichförmigen Geräuschen. *Acustica*, 1955, **5**, 303–316.
42. ZWICKER, E., & KAISER, W. Der Verlauf der Modulationsschwellen in der Hörfläche. *Acustica* 2, *Akust. Beih.*, 1952, **4**, AB 239–246.

Received January 2, 1958.

PSYCHOLOGICAL BULLETIN
VOL. 55, NO. 4, 1958

PSYCHOLOGICAL INVESTIGATIONS OF COGNITIVE DEFICIT IN ELDERLY PSYCHIATRIC PATIENTS[1]

JAMES INGLIS
Institute of Psychiatry, Maudsley Hospital, London

The view has long been accepted (**88**) that disturbances of two "components" of cognitive functions are of especial importance in the psychiatric disorders of later life. Attention has been directed to the abnormal falling away of general intelligence from a previously higher level of efficiency, and to a failure of memory, both of which seem to take place to a pathological degree in some elderly persons.

In the present paper the evidence from objective studies of these functions in elderly psychiatric patients will be reviewed. It is hoped that this review will provide a useful supplement to previous reviews of similar topics (cf. Eysenck [**29**], Granick [**34**], Grewel [**36**], Dörken [**24**], and Jones and Kaplan [**47**]) by covering the more recent literature, by trying to ascertain if any consistent relationships can be discussed within the experimental findings, and by attempting to select some framework of psychological theory which might be able to comprehend such relations as do emerge.

DECLINE OF GENERAL INTELLECTUAL FUNCTION

The studies to be examined here may be discussed under two heads: (*a*) studies mainly concerned with comparisons of *tests*, (*b*) studies mainly concerned with comparisons of *groups*.

Test comparisons. Babcock (**3**) was the first systematically to attempt to measure intellectual deficit comparing a score on a vocabulary test, supposing this to be relatively impervious to decline and therefore to represent a "high-water mark" of intelligence, with scores on other kinds of tests more affected by the process of deterioration.

Brody (**11**) adapted Babcock's test for use with elderly psychiatric patients. In one study he tested 83 patients (age range, 50–69) who fell into four "categories" of severity of dementia, A to D, D being the most severe; these being defined by clinical ratings made by doctors and nurses. He was able to show significant differences between the "discrepancy scores" of these groups and went on to suggest that this test might be used diagnostically. Those patients making a score lower than 25 would be classified, he suggested, as "probably not or doubtfully demented," those scoring over 40 would be "probably seriously demented," those between 25 and 40 being "mildly or moderately demented." However, inspection of this author's tables (**11**, p. 321) reveals the range of scores for his criterion groups (assuming normality of distribution) shown in Table 1.

The same author (Brody [**12**]) tried to see if there were a "psychometric pattern" characteristic of dementia, using Babcock's tests and others. He noted that:

> The only positive conclusion suggested by study of the psychometric pattern in the pres-

[1] The writer wishes to acknowledge suggestions and criticisms made by M. B. Shapiro and F. Post. H. G. Jones and V. Meyer also read the manuscript and offered very helpful comments.

TABLE 1

THE RANGES OF BRODY'S "DISCREPANCY SCORES" CALCULATED FROM MEANS AND *SD*s.

Group	$M \pm 2SD$	Range
A	$16.10 \pm 2(11.67)$	−7.24 to 39.44
B+*C*	$27.19 \pm 2(11.26)$	4.67 to 49.71
D	$46.25 \pm 2(13.24)$	19.77 to 72.73

ent patients is that while, in dementia, vocabulary ability is comparatively well preserved, other abilities severely decline, levelling down in a fashion which obscures the common pattern in simple psychosis and normal senility. Otherwise there is no trace of a specific pattern of abilities in dementia (**12**, p. 519).

A somewhat similar approach was adopted by Halstead (**38**) who gave a battery of 25 items to 20 senile patients (age range, 68 to 83). He was able to divide these tests into 3 groups in terms of the difficulty these patients showed in coping with them. Least difficult were those involving old mental habits, visual recognition and simple motor tasks. Vocabulary, simple arithmetic, rote memory and fluency came next. The most difficult tests were those which required the subjects to break away from old mental habits. When, however, Halstead (**39**) gave the same tests to 18 further patients (age range, 70 to 83) judged clinically to be more demented, they were shown to be worse than the first group on nearly every test, including the vocabulary scale.

Such a finding vitiates the use of vocabulary as a measure of "previous level," at least in individual cases. Theoretical weaknesses in the use of this kind of measure have previously been pointed out by Yacorzynski (**80**), and others have been examined more recently by Yates (**81**). Empirical evidence concerning the usefulness of vocabulary estimates in this connection have been provided by other workers. Trueblood (**72**), for example, found deterioration of language in 25 senile patients (ages not given). Shakow, Dolkart and Goldman (**65**) found vocabulary performance in 56 arteriosclerotics and 48 senile psychotics (age range, about 55 to 85) more disturbed than could be accounted for on the basis of age alone. Ackelsberg (**2**) tested 50 senile patients (age range, 60 to 85) grouped as "least," "mildly," and "most" deteriorated. The factors of age, education, etc., were controlled as well as possible between the groups. Five different kinds of vocabulary test were used and the results showed that vocabulary functioning did *not* remain stable but, on the contrary, discriminated significantly between these groups of patients. Roth and Hopkins (**64**) reported that 14 senile psychotics (age range, 67 to 85) whom they tested were largely inferior on a vocabulary test to a group of 84 affective psychotics (age range, 60 to 86). Orme (**55**) has shown a group of 25 senile dements (mean age, 73.84) to be significantly worse on a vocabulary test than a group of 25 elderly depressive patients (mean age, 68.17). There was, however, also a significant difference between the mean ages of the groups in the latter study. Pichot (**57**) suggested, however, that verbal deterioration may take place to a more marked degree in cerebral arteriosclerosis than in senile dementia. He showed that when 27 arteriosclerotics (mean age, 65.22) were matched with 25 senile dements (mean age, 73.08) in terms of general intellectual status on Raven's Progressive Matrices (**61**) the arteriosclerotic patients had a very significantly poorer Vocabulary score.

It may be concluded, therefore, that while it is possible that verbal

ability may decline more slowly than certain other abilities there can be little doubt that it, too, can be affected by the processes of deterioration.

M. D. Eysenck (**28**) analysed the performance of 100 clinically diagnosed senile dements (mean age, 73, *SD*, 6.5) on the Matrices test. She wished to discover if these patients differed from the normal standardization population not only in terms of a lower total score but also in terms of qualitative characteristics such as order of difficulty of items within the test and most frequent errors made. She was able to show that while these patients made low scores, approximately equal to the lowest 3 or 4% of the normal adult population on whom the test was standardized, they were not, in fact, qualitatively different from children or normal adults in the respects measured. Bromley (**13**) similarly examined the results on the Matrices test of 35 elderly psychiatric patients (mean age, 61, *SD*, 11.3) including 8 diagnosed as senile, 10 paranoid states, 12 depressives, and 5 organics. His results agreed, in the main, with those obtained by Eysenck, in showing a diminished total quantitative score, but little qualitative difference from normals.

Eysenck (**27**) also factor analysed the results of a battery of 20 tests, including the Matrices, given to 84 senile patients (mean age 73.4, *SD*, 6.5). She extracted one general and 3 group factors (speed, memory, and strength) accounting for 43% of the variance in all. Her main conclusion from this analysis was that the general factor extracted gave a different picture of the mental organization of such patients as compared with normal adults. The test with the highest saturation on the general factor was Vocabulary (.71) whereas the Matrices showed a low saturation (.34). Eysenck attempted to explain these results in terms of Cattell's (**21**) notions regarding "fluid" and "crystallized" intelligence. She concluded that, on the basis of these conceptions:

> . . . we should expect tests of fluid ability, such as the matrix, to have very high factor saturations in a study of adolescents and young people generally but to give low factor saturations in extreme old age; while, conversely, crystallized abilities would give comparatively high factor saturations in old age. This is precisely what we do find in this analysis, and we may, therefore consider that this analysis gives a certain measure of support to the theory set forth by Cattell (**27**, p. 18).

Pinkerton and Kelly (**58**), also used the Matrices and examined the results of 40 patients (age range, 60 to 90) clinically graded into five groups from A to E, E being the most demented. These authors attempted to evaluate not only extent of intellectual deterioration but also, "the degree of emotional adjustment to that deterioration." In their study, quantitative scoring showed significant differences only between groups A and E. Their qualitative, emotional judgment score involved principally the kind of error made by patients after their last correct choice. It was suggested that: "By determining the ratio between the percentage marks up to the limit of problem difficulty reached, and the percentage score in the problems beyond that point, a rough index of emotional adjustment to deterioration is given" (**58**, p. 251). It seems to the writer that while it might be of interest to obtain further external evidence concerning the meaningfulness of such a score, merely to have these authors' demonstration that on the whole, the *more* demented obtain a *better* score on emotional adjustment, makes one rather sceptical of its "reality value."

Cleveland and Dysinger (**22**) tested 20 senile patients (age range 64 to 83) on the Wechsler-Bellevue Intelligence Scale (**74**) and on the Goldstein-Scheerer Object Sorting Test (**33**). They showed that these patients tended to experience marked difficulty on the Performance Scale of the Wechsler and that this difficulty was commonly associated with a complete inability to handle the kind of abstractions required for success on the sorting test, even if the patient were of fairly average ability on the Wechsler Verbal Scale.

Lovett Doust et al. (**52**) also used the Wechsler scale in a study designed to examine the relations between psychometric status and arterial oxygen saturation (measured spectroscopically). The subjects were 83 patients (age range, 62 to 88) so selected clinically as to represent a "continuum of senile dementia." A highly significant correlation ($r = .47$) was found between the psychological and physiological indices. It is of interest to note that the mean subtest intercorrelation ($r = .63$) for this abnormal group was higher than that reported for the normal population on whom the test was standardized.

Lassen et al. (**50**) also showed a relationship between the clinical estimate of dementia, cerebral oxygen consumption and performance in 19 patients (18 of whom were between 42 and 67) on the Goldstein Scheerer Block Design test.

Silverman et al. (**67**) in one of a series of studies which included 56 hospitalized senile patients (all over 60) showed a relation between intellectual efficiency and EEG findings such that elderly persons with "mixed" or "diffuse" EEGs did worse on psychological tests, including the Wechsler, than those with "normal" or "focal" EEGs. These authors also showed that the "mixed" or "diffuse" picture predominated in the hospitalized senile patients.

Group comparisons. A comparative study involving several groups of elderly psychiatric patients is reported by Rabin (**59**). He compared the Wechsler results of four groups, clinically diagnosed as senile ($N = 15$), arteriosclerotic ($N = 30$), miscellaneous ($N = 30$), and nonpsychotic ($N = 25$). He failed to find significant differences in subtest patterning between the groups, but the total quantitative scores were lower in the senile and cerebral arteriosclerotic groups. The Verbal Performance discrepancies also varied in size, being largest in the seniles, next largest in the cerebral arteriosclerotics, less in the miscellaneous group and least in the nonpsychotics. It should be remarked, however, that Rabin reported differences in mean age between the groups, although not enough data are presented to enable the significance of this difference to be checked. Rabin himself concluded that age might have been the relevant factor which produced the found differences in performance.

Botwinick and Birren (**9**) examined the validity of three separate indices sometimes used to detect abnormal intellectual deterioration in the elderly. These were, (*a*) the Deterioration Quotient (DQ) derived from the Wechsler-Bellevue by comparing "hold" against "don't hold" subtests, (*b*) the Efficiency Index (EI) from the Babcock test, and (*c*) the Senescent Decline Formula devised by Copple (**23**) and based on the Wechsler. The experimental group comprised 31 hospitalized patients clinically diagnosed as psychosis with cerebral arteriosclerosis or senile psychosis (age range, 60 to 70). The control group consisted of 50

normal elderly persons (age range, 60 to 69) (cf. Fox and Birren [**31**]). Significant differences were found between the patients and the controls on the EI and SDF measures, but not on the DQ measure. These authors concluded that:

> The measure which best differentiated the two groups was the EI. This index contains a category called "initial learning" which is comprised of three subtests. The difference between the two groups for this category was greater than the differences in any of the Wechsler-Bellevue subtests. Future attempts to refine measures of intellectual deficit might well exploit the use of such learning subtests (**31**, p. 148).

In another paper Botwinick and Birren (**10**) compared the same patients and controls in terms of their differential performance on each of the Wechsler-Bellevue subtests. They were able to show that the amount of decline in any subtest with age was not necessarily a criterion of that subtest's ability to differentiate between the experimental and control groups.

Birren (**7**) factor analysed the Wechsler results of these 31 patients and also the results of a group of 99 normal elderly persons (age range, 60 to 74). Four factors were extracted from the correlation matrices; these were labelled Verbal Comprehension, Closure, Rote Memory and Induction. Before this analysis took place, however, it was found that the median correlation between subtests for the abnormal ($r = .63$) was higher than that for the normal group ($r = .53$). This is in close agreement with the higher intercorrelation for an abnormal group found by Lovett Doust et al. (**52**).

Dörken and Greenbloom (**25**) compared the Wechsler results of 67 patients (age range, 66 to 89) with a clinical diagnosis of either senile psychosis or cerebral arteriosclerosis with the results of a control group of 20 normal subjects (age range, 65 to 80). They found that the abnormal group were lower in intelligence, showed larger discrepancies between the Verbal and Performance scale results and also had higher intercorrelations between their subtest scores. They also found Copple's SDF to be a more valid measure than Wechsler's DQ.

Robertson and Wibberley (**63**) gave a large battery of tests to 24 housewives (age range, 40 to 60); eight of these were clinically stated to be "demented," eight were dull but not deteriorated and eight were defective but not deteriorated. The original ability of these patients was very carefully established by biographical enquiry and the independent ratings of several judges. It was found that the original ability of the demented group was best reflected in tests involving verbal material while their deterioration showed itself in material involving visuo-spatial ability.

Roth and Hopkins (**64**) compared the results of 14 senile psychotics (median age 78) with those of 46 elderly "functional" patients (median age 67) on four tests, Wechsler Vocabulary, Digit Span, Raven's Matrices and an Information test. Since these groups differed in age, a further control group from a general hospital, comprising 14 elderly patients of the same age as the senile group, was also tested. Significant differences were found between the functional and normal subjects on the one hand and the senile psychotics on the other on all four tests. However, while the ability of the seniles did not overlap with the other groups at all, for example, on the Matrices, their ability to define words and repeat digits was relatively well preserved

in some cases. Hopkins and Roth (**42**) compared the above results with those of further groups of elderly psychiatric patients diagnosed as paraphrenic ($N=15$), arteriosclerotic psychosis ($N=22$), and as acute confusional state ($N=17$); there was no significant difference in age between these groups. Test scores placed the paraphrenics and confusional states with the functional patients, while the arteriosclerotics overlapped both with these and the senile group.

Orme (**55**) examined the results of 25 clinically diagnosed cases of senile dementia (mean age, 73.84) and a group of 24 elderly depressed patients (mean age, 68.17) and two groups of healthy old people (**56**) on the Matrices and the Mill Hill Vocabulary Scale (Raven [**60**]). Significant differences were shown between these groups on both tests; they were, however, also significantly different in mean age. Orme claims that the differences found between the tests cannot be accounted for in terms of the age difference, since the only significant correlation with age is between it and the vocabulary scale within the senile group alone. He also suggests, contrary to the usual beliefs about intellectual deterioration, that a decline in *verbal* ability may be the most fundamental characteristic. This latter argument he supports by showing the groups to be less different on a score of "intellectual potential" (derived from the Matrices by taking as a score the last point where a particular problem was solved on two occasions of testing) than on a score of "intellectual productivity" (or the simple sum of items correct on the test). "By this method," the author states, "it is possible to assess each patient's maximum potential of intellectual ability disregarding losses during performance caused by an apparent fluctuation in ability" (**55**, p. 865). Whatever the merits of this measure, it seems to the present author that Orme's contention is somewhat weakened by the fact that only within the depressed group are there significant correlations between either "potential" and the Vocabulary score or "productivity" and the Vocabulary score. If, as Orme claimed, difference in verbal ability determined the disturbance in intellectual productivity, then it seems that significant correlations would be expected to appear between at least the "productivity" measure and the Vocabulary score *within* the senile group itself.

In addition to the comparative work reported above, which has mainly used fairly well-standardized tests, a few studies have been reported which have employed less well-standardized material for the examination of cognitive functions.

Thus, Hall (**37**) studied 70 patients (age range, 41 to 65) categorized as "definitely organic," "definitely depressive," or "doubtful" on the basis of physical, CSF, AEG and EEG examination. He found that conceptual tests, including a block sorting test, differentiated best between the groups.

On the other hand Hopkins and Post (**41**) showed that while 49 psychiatric patients (age range, 60 to 84) were, on the whole, less capable on such tasks as the Goldstein-Scheerer Cubes and Colour Form Sorting Test than 49 normal controls (age range, 60 to 87), there were no striking differences within subclasses of the psychiatric group. They therefore conclude that:

> The differentiating value of the Goldstein tests used appears to be limited to the negative finding that preservation of the abstract attitude in elderly people makes the presence

of organic cerebral disorder unlikely. Failure In test performance does not necessarily indicate the presence of brain damage (**41**, p. 849).

Thaler (**71**) also found that 66% of her group of 116 normal old people (mean age, 73.04; *SD*, 7.53) performed concretely on the Colour Form Sorting test.

Studies of more specific aspects of mental function in elderly psychiatric patients include those of Birren and Botwinick (**8**) who examined the effects of age and senile psychosis on motor speed (in writing digits and words). They contrasted the results of 35 patients (age range, 60 to 70) institutionalized for senile psychosis or cerebral arteriosclerosis with psychosis with the results of a control group of normal persons of the same age. Statistically significant differences were found in writing speed, the patients being slower than the normals.

Williams (**77**) has studied the differences in responses made to visual stimuli by groups of patients diagnosed as suffering from early senile dementia and mentally normal elderly persons. The former groups performed more poorly, only improving if the tasks were simplified or if additional cues were presented.

Discussion. The main trends which emerge from the review of studies of decline in general intellectual functioning seem to be as follows.

1. Given the clinical diagnosis of dementia, psychometric studies in general show patients so categorized to be poorer intellectually than control groups of other elderly persons, whether "functional psychotics" or normals. In a way, this is hardly surprising since one of the main criteria commonly used for placing a patient in the "dement" category is a clinical impression of cognitive impairment. Such a process of indirect contamination has been studied by Shapiro et al. (**66**) and has been shown to be significant. However, this latter process is only of major importance if a hypothesis is being tested, e.g., that such deficit is due to organic aetiology. It may be of less importance when the purpose of the study is mainly descriptive.

2. Some studies have, however, shown a relation between psychological adequacy and physiological efficiency, confirming that the maintenance of general intellectual functioning may depend to some degree on cerebral metabolism, especially on the cerebral uptake of oxygen (**50**, **52**).

3. Some studies have demonstrated that higher intercorrelations may be found in intellectual test performance among abnormal groups (**7**, **52**).

4. The existence of some kind of "verbal-performance discrepancy" in cases of senile deterioration has been found in a number of studies (**10**, **11**, **22**, **25**, **38**, **59**, **63**). Other studies, however, have made the important point that verbal ability can by no means be regarded as an infallible, constant "intellectual high-water mark" and have shown that verbal ability may itself be affected by the process of deterioration (**2**, **39**, **55**, **57**, **64**, **65**, **72**).

5. The relation of such findings as these latter to differences between "fluid" and "crystallized" ability and the relation of these, in turn, to new and old learning has been emphasized by Eysenck (**27**). Birren also has noted that:

> Intelligence tests such as the Wechsler-Bellevue seem to measure some diffuse combination of (*a*) the mass of previously learned or stored information and (*b*) abilities to acquire information. Success on a test item would seem to be determined by whether a similar

bit of information has previously been presented to the individual, by his capacity to store it, and by his ability to associate a currently presented bit of information with relevant stored information and pertinent information in the context of the item. As an individual grows older his environment changes in the relative frequency of occurrence of certain types of information. In addition to the age changes in frequency with which certain information occurs or is reinforced, there is the possibility that the individual suffers a diminution in abilities. He may not learn as rapidly or associate a presented bit of information with his stored information. . . . The close association of test performance in the elderly with stored information suggests that to a greater extent than in young subjects their performance is determined by what they already know than by what new information they can elicit from the test situation. With senile brain changes there is not only a further reduction in the ability to extract new information but there is also a large loss of stored information as well (**7**, pp. 403–405).

Psychological studies point, in fact, to a connection between intellectual deterioration and some disorder in learning ability. It remains to be seen whether evidence for such a connection between these two aspects of function emerges, so to speak, from the other side, in studies of "memory disorder" itself.

Memory Disorder

Before psychological studies in this area are examined it is necessary to emphasize two considerations, one mainly theoretical, the other more practical, both concerned with the notion of memory in general.

In the first place, neither the status of "memory function" as a construct, nor its relation to other cognitive abilities has ever been clearly decided. Clinically it has commonly been considered as distinct from general intellectual functioning but as Spearman (**69**) emphasized many years ago, clinical usage need not always be an ideal guide in such matters. Simmins (**68**), for example, found in a sample of 200 mental hospital patients that when *g* scores were partialled out from results on so-called memory tests, the ability of over 90% of her patients fell within the normal range. Eysenck and Halstead (**26**) showed in a normal group of 60 young soldiers that most of the variance on several "memory tests" could be accounted for in terms of general intelligence.

However, a study by Inglis et al. (**46**) has suggested that the notion of "memory function" may be a useful one in the study of the behaviour of elderly psychiatric patients even when general intellectual level has been partialled out of their performance.

Secondly, as Hull (**43**) pointed out more than 40 years ago, the notion of "memory" as commonly used, is too narrow in meaning, having come to refer almost exclusively to the process of reproduction of learned material. It might be better to substitute "learning ability" for memory even in clinical discussions, since the former expression is wider in connotation. It admits of several aspects and takes into account, for example, at least the two broad phases of "fixation" or "acquisition," and "retention." So long as it is clear that the distinction between these phases is not absolute, as McGeoch and Irion (**53**) have pointed out, it is likely that "learning" will be a more convenient term than "memory" for the analysis of the behaviour concerned.

Bearing these considerations in mind, studies which have attempted experimentally to examine learning ability in elderly psychiatric patients may be examined in chronological order.

The results of the earliest studies can, at best, be regarded as only suggestive. They commonly employed

very few, and very mixed cases, they usually omitted to control important independent variables and seldom carried out any statistical checks.

Among such studies is a report by Achilles (**1**) who examined, among other subjects, 16 mixed, mainly elderly, psychiatric cases (age range, about 36 to 84), suffering from Korsakoff's syndrome, arteriosclerosis, and GPI. The subjects were not required to learn the material used up to a set criterion but were simply given 2 presentations of two sets of words and one set each of "forms," "proverbs," and "syllables." Achilles concluded that, "No attempt will be made . . . to average the results of the subjects in the insane group but . . . all show a memory defect and the defect is present in both recall and recognition" (**1**, p. 64).

Moore (**54**) reported a study of 30 patients (age range, 28 to 74, 14 of whom were over 50) including cases suffering from GPI, senile dementia and alcoholic deterioration. These subjects were presented once with 8 stimuli in each of 4 series, real objects, pictures of objects, and printed and spoken words. These had to be recalled immediately and then again one minute later. The patients were also required to name 50 pictures. Moore found evidence suggesting that, while immediate memory, retention, and perception all tend to deteriorate together, nevertheless, immediate memory and retention represent "two distinct mental functions."

Liljencrants (**51**) examined only 4 abnormal cases (ages 40, 53, 54, and 60) and concluded that both apprehension and reproduction are affected in cases of memory disorder when the patients are required to recall or recognize meaningful or nonmeaningful visual stimuli.

Wylie (**79**) studied seven senile, three presenile, and six paretic, "deteriorated" cases (age range, 29 to 87, 11 of whom were over 50). Her results suggested that such patients could neither acquire nor retain in situations in which they were required to remember pictures or to learn paired associates, including meaningful, nonmeaningful and nonsense material. They failed both in the recognition and recall of such material.

Hunt (**44**) noted in his review of such studies that: "The control of most of the psychological governors is inadequate because it is not understood. Even the conception of memory varies. Nevertheless, certain facts stand out" (**44**, p. 13). Two of these outstanding facts are of importance here: (*a*) "impression" often suffers more than retention, and (*b*) inefficiency is evident in tests of both recall and recognition.

An example of a more sophisticated approach to the study of memory disorder in elderly psychiatric patients is provided by the work of Shakow, Dolkart, and Goldman (**65**). They used a version of the Wells Memory Test (Wells and Martin [**76**]), which involves both old recall (e.g., of personal information) and new recall (e.g., memory for a sentence). They examined 56 cerebral arteriosclerotics with psychosis (age range, approximately 55 to 75) and 48 senile psychotics (age range, approximately 56 to 85). The performance of these groups was compared with that of normal controls selected from the same age range of 192 normals (age range, 15 to 90) also tested. The only control for general level was in terms of previous education since the authors found both abnormal groups were lower in Vocabulary score than the normals. They found consist-

ently, although not markedly, poorer results on these memory tests in the psychotic groups. All groups were better on old than on new recall and seniles were poorer than arteriosclerotics on both tests. The authors do not, however, provide assessments of the statistical significance of these differences. It is of interest that in this study also there were higher intercorrelations between tests in the senile group than in the normals or the arteriosclerotics.

D. E. Cameron also conducted a series of investigations into "memory disorder" in aged patients. In one study (**17**) he tested 33 elderly patients (ages not given) showing memory defect. Eighteen of these were given pairs of drawings to copy and also to draw from recall. Cameron found great "interference" between the pairs and suggested that this was due to a greater tendency to perseverate. Fifteen other patients were asked to name two simple objects which had been presented on three successive occasions. They showed a tendency to "secondary elaboration" of the material. The author concludes that,

> . . . in patients suffering from the psychoses of the senium there are a greatly increased tendency to perseverate and a greatly accelerated tendency to secondary elaboration of memorized data. The first process, by interfering primarily with registration, and the second process, by interfering with retention, contribute materially to the impoverishment of recent memory in these patients (**17**, p. 992).

The same investigator (Cameron [**18**]) showed that, in 16 elderly patients (ages not given) suffering nocturnal delirium, their wandering and confusion could be brought on by their being placed in a darkened room during the day. This demonstrated that their condition was not brought on by fatigue, as has sometimes been supposed, but by cue-deprivation. Since these patients also showed severe learning disability Cameron suggested that such delirium may be based on an inability to maintain a spatial image without the assistance of repeated visual stimulation. Thirteen of these patients also showed a severe distortion of spatial imagery (e.g., in the displacement of the remembered positioning of environmental objects) within an hour of being blindfolded.

Cameron (**19**) also compared 12 seniles (ages not given) with memory defect with a group of younger persons on a test of memorizing 3 digits. These they were required to reproduce after varying periods of time (cf. Feldman and Cameron [**30**]). If the interval between acquisition and reproduction was unfilled, the seniles could remember the series for brief periods. If the period was occupied by another task, however, they failed in recall, unlike the controls who showed no such difficulty.

In all, Cameron gives emphasis to the importance of disturbances of the *early* stages of retention, or what the present author would prefer to call the "acquisition" stage in learning.

Cameron et al. (**20**) also showed in 23 patients (age range, 60 to 87) suffering from psychoses of the senium characterized by "memory defect," that they also suffered defects of cerebral oxygen metabolism.

A study by Kral and Durost (**48**) using the Wechsler Memory Scale (**75**) included 10 senile dements (mean age, 78) and 10 Korsakoff's syndrome (mean age, 51.5). They found impairment of immediate recall and recent memory in both these groups. The control group in this study was inadequate, comprising 10 hospital personnel with a mean age of only 29.

Williams (**78**) demonstrated in a study of 60 senile dements (age range, 65 to 95) that such patients may experience gross difficulties in learning even the simplest paper-maze tasks, although their performance can be aided to some degree by recent performance on similar tasks. No control data are quoted in this study.

At least one further investigation has demonstrated a relation between learning ability and physiological efficiency in relation to the oxygen uptake of the brain. Lassen et al. (**50**) using 19 patients (18 of whom were between 42 and 67) found a relationship between the clinical impression of dementia, cerebral oxygen metabolic rate, and the ability to learn a series of digits when this was only slightly longer than the immediate memory span. These authors reported the results on this test in terms of a "fluctuation score." They found *least* fluctuation in the nondemented *and* the most demented groups, fluctuation appearing only in the middle range. It seems to the present writer that this finding may have been due to the fact that the test was too easy for one extreme group and too difficult for the other, variations only being possible in the middle group. It is likely that a more meaningful score would have been obtained if they had counted the number of trials to the criterion of learning. Zangwill (**82**) previously found this method to provide a useful test of learning in organic cases.

Robertson (**62**) has recently reported an experiment in which, among other subgroups, 26 elderly patients (age range, 60 to 79) diagnosed as suffering from various kinds of brain damage including "organic senile dementia" were compared with 40 non-brain-damaged, psychotic patients (age range, 60 to 79). The experimental group were found to be significantly worse on tests of paired associate learning. When both groups were further subdivided into a "low vocabulary" (Wechsler Vocabulary score, less than 23) and a "high vocabulary" group (23 or above), it was shown that the low vocabulary groups were also significantly different on a rote-learning test, and the high vocabulary groups almost so. This study clearly demonstrated the need for controlling at least the factors of verbal ability and age in any experimental study of learning ability.

Discussion. From this review of the experimental work concerned with disturbances of learning ability in elderly psychiatric patients, it may be seen that there are certain findings or features in common with the studies of intellectual deterioration already reviewed.

1. Given the clinical diagnosis of dementia, patients so categorized can be shown to suffer some disabilities in learning, both in acquisition and in reproduction (recognition and recall).

Many of the studies reviewed have, however, involved poor experimental design and have failed to adhere to adequate experimental techniques (e.g., they have not studied learning through the various sense modalities, nor have they set up adequate criteria of learning).

In addition, most of these investigations are affected by the same process of "indirect contamination" as were the studies of intellectual deterioration. This is of considerable importance if it is intended to test the common psychiatric hypothesis concerning the "organic" aetiology of memory disorder, or if it is required to show that there is some necessary connection between intellec-

tual deterioration and learning disorder in senility.

2. Some studies have, however, suggested that learning efficiency may be related to physiological adequacy. Indirectly this notion is supported by some of the findings made by Simmins (**68**) and also by Inglis et al. (**46**). More direct evidence comes from those studies which have related memory function to cerebral oxygen consumption (**20, 50**).

3. At least one study has confirmed the existence of higher correlations between learning and other tests in disordered groups (**65**).

4. That some important relationship exists between some aspects of general intellectual deterioration and learning disability is confirmed by a recent study by the present author (**45**). Using small groups of elderly psychiatric patients (age range, 56 to 81) *selected solely on the basis of the presence or absence of clinically ascertained memory disorder*, and matched for age and verbal ability, he was able to demonstrate that: (*a*) Experimental and control groups did *not* differ in terms of immediate memory span for digits forwards. (*b*) They did, however, differ very significantly in their rate of and degree of success in the acquisition of paired associate items, whether visually or auditorily presented, whether recalled or recognized. (*c*) The experimental group showed a significantly larger Verbal-Performance discrepancy on the Wechsler than the control group, and (*d*) There was a higher correlation between such a discrepancy and learning (i.e., the acquisition phase) in the experimental than in the control group.

Psychological Evidence from Attempts at Treatment

In addition to those studies already considered, in which attempts were made to examine psychological functions more or less in isolation, there have been reported in the literature a number of studies which have been concerned to improve physiological and metabolic functioning in elderly psychiatric patients and to assess the effects of such treatment on psychological function.

Three main therapeutic avenues have been explored; these comprise treatment through increase of oxygen utilization, vitamin therapy, and sex hormone administration.

Cameron (**19**), for example, attempted to increase the supply of oxygen to the brain and also tried to increase the utilization of oxygen by the brain in elderly psychiatric patients, with negative results. A study by Garnett and Klingman (**32**) used injections of an oxygen-transferring enzyme, Cytochrome C, in 17 mainly elderly psychiatric patients (age range, 43 to 79, 11 of whom were senile, presenile or arteriosclerotic). These authors claimed "clinical improvement" of memory and other functions in most of these cases. There was, however, no significant change in Wechsler scores between pre- and posttreatment assessment. No untreated control group was used in the study.

Cameron (**19**) also administered vitamins (B_1, nicotinic acid, etc.) without result. Stephenson et al. (**70**) had previously shown that vitamin treatment had beneficial effects at least on psychomotor performance in 40 senile dementia patients (age range, 65 to 86) as compared with 18 untreated senile patients (age range, 60 to 87). The outcome of a study of vitamin administration in a large group of male senile patients reported by Vernon and McKinley (**78**) was, however, almost wholly negative

both as regards psychomotor and intellectual test performance. On the other hand a paper by Gregory (**35**) concluded that vitamin therapy is effective in senile disorders of recent onset. This author administered nicotinic acid to his experimental group but used no untreated controls. He reported only the results of his successful cases and in any case his estimates of improvement were mainly based on clinical impressions. Although intelligence tests were used in some cases the tests were not described. In view of these facts Gregory's conclusions must be regarded as speculative. A well designed study has recently been reported by Krawiecki et al. (**49**). These authors demonstrated improvement on the Wechsler Memory Scale in 25 senile psychotics (age range, 65 to 83) following injections of the vitamin preparation, Parentrovite, as compared with 25 placebo-treated senile controls matched for age and pretreatment memory ability.

The effect of sex hormone administration on psychological functioning was also examined by Vernon and McKinley (**78**). Using male hormone in their male senile patients they found no effect on either psychomotor or intellectual tests. The results of a well-designed and controlled study conducted by Caldwell and Watson (**15**), however, showed that, in nonpsychotic cases at least, the administration of female sex hormones to 15 elderly women (age range, 54 to 88) improved their performance on the Wechsler Memory Scale, especially on the Paired Associates subtest. This improvement was significantly greater than that of a matched, placebo-treated control group. Further reports by Caldwell and Watson (**16**) and Caldwell (**14**) showed that the experimental group continued its improvement over at least a year, so long as the hormone administrations were continued.

Such studies serve to confirm the relation already discerned from studies principally concerned with intellect and learning, that some relation exists between physiological, probably metabolic, adequacy and psychological efficiency in the psychiatric disorders of old age.

Theoretical Considerations

The view has often been expressed that human cognitive behaviour is to be understood in terms of some balance between the capacity to acquire and the ability to use what has been acquired. Thus we have the notions of "fluid" and "crystallized" ability proposed by Cattell (**21**), and the processes of "accommodation" and "assimilation" suggested by Piaget (Berlyne [**5**]). One outstanding attempt functionally to relate these processes to each other and to unite them within a theory of brain function has been made by Hebb (**40**).

Briefly, the "neuropsychological" theory propounded by Hebb suggests that short-term "memory" may be maintained by rather complex reverberating neuronal circuits, while repeated excitation of these would lead to permanent changes in the interconnections of their constituent neurones, so forming the "cell-assemblies" which are proposed as the basis of learning (and so of longer lasting "memory" processes). Activity of such circuits can be facilitated from sensory input or from other assemblies; this latter process can, in turn, result in the creation of more or less permanent processes of interfacilitation between assemblies to form "phase-sequences," which may be considered to provide the neural basis of "attention" and even of "thought."

This theory provides at least a speculative basis for regarding learning and intelligence as lying along some "physiological continuum." Intelligence itself Hebb (**40**, p. 275f.) regards as dividing into "intelligence A," or innate potential, and "intelligence B," the functioning of the brain in which development and learning have gone on.

Integration of this theory with the experimental findings reviewed above may first be pursued through the psycho-physiological relations discovered by some investigators (**20**, **50**, **52**, **67**). Hebb has pointed out that:

> A (brain) region in which the blood supply is interfered with, but not entirely shut off, usually shows some loss of cells and a number of remaining cells whose staining properties are changed. This indicates a change in the chemical properties of the cell, which in turn implies a change in frequency properties and obviously may account for the existence of a hypersynchrony which interferes with the functioning of the cell-assembly. A focus of hypersynchrony must act as a pacemaker that tends to wean transmission units away from the assembly. When hypersynchrony is not great, it would allow some assemblies to function (particularly those that are long established) but would tend to interfere with recent memory, decrease responsiveness and interfere with complex intellectual activities. When it is more extensive, it would prevent all higher functions (**40**, p. 283).

Concerning long-established assemblies, Hebb notes that,

> One might assume that longer-established assemblies would have in general greater safety margins. . . . This would imply that older habits, and longer-established memories, would be most resistent to disruption by metabolic changes of blood content (**40**, p. 197).

It is tempting to try further to relate these speculations to those of such workers as Eysenck (**27**) and Birren (**7**) regarding the reorganization of mental abilities in the older adult.

They could be used to explain not only why Verbal-Performance type discrepancies have been found, but also why, if the disorder is far advanced, verbal ability (or "old learning") may itself decline. It may be suggested that the psychological efficiency of elderly persons mainly depends on crystallized ability but nevertheless a modicum of "fluidity," or learning ability, is also very necessary. When this latter drops below a certain "threshold" level not only is ability on learning tasks affected, but the type of discrepancy sometimes noted between Verbal and Performance tasks makes its appearance, since Performance type tests require more "learning" of instructions and present more unfamiliar problems to be solved. This hypothesis is also congruent with the results found on "sorting" tasks (**22**, **37**), which are commonly of the "concept formation" type.

The notion of the proposed "threshold effect" is also consistent with an idea put forward by Shakow et al. (**65**) who noted that, since they found no correlation *within* the senile group between age and memory, " . . . a person who develops senile psychosis at 80 is an individual who at 80 has preserved his functions as well as the person who develops the psychosis at 60. It is as if there were a threshold level of preservation of function and when that threshold is reached the person is ready to succumb to the psychosis" (**65**, p. 46).

The increased correlations between functions frequently found (**7**, **52**, **65**), in the abnormal groups might also be due to this threshold effect. For example, two tests, X and Y, may have a "minimum requirement" in terms of learning, which all "normal" persons possess, and otherwise have little variance in common. Once

the individual falls below a certain threshold of ability, however, the common, "learning" element comes to be of crucial importance and a higher correlation results.

At a later stage even crystallized ability (or "intelligence B") could be disrupted with the consequent falling off, for example, in verbal ability.

Attention may finally be drawn to a striking resemblance between the results of Cameron's (**18**) cue-deprivation experiment with senile patients and some experiments issuing from Hebb's own laboratory. Bexton et al. (**6**), working with 22 young normal volunteers, found that sensory deprivation, with its consequent diminution of sensory controlling input, over relatively long periods of time (i.e., 2 to 3 days) resulted in gross disturbances of behaviour. When, as Hebb has described, the functioning of the assemblies is itself adversely affected by metabolic changes it might be expected that much shorter periods of cue-deprivation would lead to disturbances of orientation and behavior, such as Cameron (**18**) showed in his elderly patients. Bartlet's (**4**) review of visual hallucinations in intellectually well-preserved elderly patients with cataract is of interest in this connection.

Summary

1. An attempt has been made to review the experimental literature to date on intellectual and learning impairment in elderly psychiatric patients.

2. Certain consistent relationships emerge from these studies indicative of possible links between learning ability and that differential cognitive impairment which is to be found in at least the early stages of senile deterioration.

3. The neuropsychological theory propounded by Hebb seems to provide a convenient conceptual framework for the relations so far established.

REFERENCES

1. Achilles, Edith M. Experimental studies in recall and recognition. *Arch. Psychol.*, 1917, **6**, No. 44.
2. Ackelsberg, Sylvia B. Vocabulary and mental deterioration in senile dementia. *J. abnorm. soc. Psychol.*, 1944, **39**, 393–406.
3. Babcock, Harriet. An experiment in the measurement of mental deterioration. *Arch. Psychol.*, 1930, No. 117.
4. Bartlet, J. E. A. A case of organized visual hallucinations in an old man with cataract and their relation to the phenomena of the phantom limb. *Brain*, 1951, **74**, 363–373.
5. Berlyne, D. E. Recent developments in Piaget's work. *Brit. J. educ. Psychol.*, 1957, **27**, 1–12.
6. Bexton, W. H., Heron, W., & Scott, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, **8**, 70–76.
7. Birren, J. E. A factorial analysis of the Wechsler-Bellevue scale given to an elderly population. *J. consult. Psychol.*, 1952, **16**, 399–405.
8. Birren, J. E., & Botwinick, J. The relation of writing-speed to age and to the senile psychoses. *J. consult. Psychol.*, 1951, **15**, 243–249.
9. Botwinick, J., & Birren, J. E. The measurement of intellectual decline in the senile psychoses. *J. consult. Psychol.*, 1951, **15**, 145–150.
10. Botwinick, J., & Birren, J. E. Differential decline in the Wechsler-Bellevue subtests in the senile psychoses. *J. Gerontol.*, 1951, **6**, 365–368.
11. Brody, M. B. The measurement of dementia. *J. ment. Sci.*, 1942, **88**, 317–327.
12. Brody, M. B. A psychometric study of dementia. *J. ment. Sci.*, 1942, **88**, 512–533.
13. Bromley, D. B. Primitive forms of responses to the Matrices test. *J. ment. Sci.*, 1953, **99**, 374–393.
14. Caldwell, Bettye McD. An evaluation

of psychological effects of sex hormone administration in aged women. II. Results of therapy after eighteen months. *J. Gerontol.*, 1954, **9**, 168–174.

15. CALDWELL, BETTYE MCD., & WATSON, R. I. An evaluation of psychologic effects of sex hormone administration in aged women. I. Results of therapy after six months. *J. Gerontol.*, 1952, **7**, 228–244.
16. CALDWELL, BETTYE MCD., & WATSON, R. I. An evaluation of sex hormone replacement in aged women. *J. genet. Psychol.*, 1954, **85**, 181–200.
17. CAMERON, D. E. Certain aspects of defects of recent memory occurring in psychoses of the senium. *Arch. neurol. Psychiat.*, 1940, **43**, 987–992.
18. CAMERON, D. E. Studies in senile nocturnal delirium. *Psychiat. Quart.*, 1941, **15**, 47–53.
19. CAMERON, D. E. Impairment of the retention phase of remembering. *Psychiat. Quart.*, 1943, **17**, 395–404.
20. CAMERON, D. E., HIMWICH, H. E., ROSEN, S. R., & FAZEKAS, J. Oxygen consumption in the psychoses of the senium. *Amer. J. Psychiat.*, 1940, **97**, 566–572.
21. CATTELL, R. B. The measurement of adult intelligence. *Psychol. Bull.*, 1943, **40**, 153–192.
22. CLEVELAND, S., & DYSINGER, D. Mental deterioration in senile psychosis. *J. abnorm. soc. Psychol.*, 1944, **39**, 368–372.
23. COPPLE, G. E. *Senescent decline on the Wechsler-Bellevue Intelligence Scale.* Unpublished doctoral dissertation, Univer. Pittsburgh: 1948.
24. DÖRKEN, H. Psychometric differences between senile dementia and normal senescent decline. *Canad. J. Psychol.*, 1954, **8**, 187–194.
25. DÖRKEN, H., & GREENBLOOM, GRACE C. Psychological investigations of senile dementia. II. The Wechsler-Bellevue Adult Intelligence Scale. *Geriatrics*, 1953, **8**, 324–333.
26. EYSENCK, H. J., & HALSTEAD, H. The memory function. I. A factorial study of fifteen clinical tests. *Amer. J. Psychiat.*, 1945, **102**, 174–180.
27. EYSENCK, M. D. An exploratory study of mental organization in senility. *J. neurol. neurosurg. Psychiat.*, 1945, **8**, 15–22.
28. EYSENCK, M. D. A study of certain qualitative aspects of problem solving behaviour in senile dementia patients. *J. ment. Sci.*, 1945, **91**, 337–345.
29. EYSENCK, M. D. The psychological aspects of ageing and senility. *J. ment. Sci.*, 1946, **92**, 171–181.
30. FELDMAN, F., & CAMERON, D. E. The measurement of remembering. *Amer. J. Psychiat.*, 1944, **100**, 788–791.
31. FOX, CHARLOTTE, & BIRREN, J. E. Intellectual deterioration in the aged: agreement between the Wechsler-Bellevue and the Babcock-Levy. *J. consult. Psychol.*, 1950, **14**, 305–310.
32. GARNETT, R. W., & KLINGMAN, W. O. Cytochrome C: effects of intravenous administration in presenile, senile and arteriosclerotic cerebral states. *Amer. J. Psychiat.*, 1949, **106**, 697–702.
33. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, 239.
34. GRANICK, S. Studies of psychopathology in later maturity—a review. *J. Gerontol.*, 1950, **5**, 361–369.
35. GREGORY, I. Nicotinic acid therapy in psychoses of senility. *Amer. J. Psychiat.*, 1951, **108**, 888–895.
36. GREWEL, F. Testing psychology of dementias. *Folia. psychiat. Neur.*, 1953, **56**, 305–339.
37. HALL, K. R. L. Conceptual impairment in depressive and organic patients of the pre-senile age group. *J. ment. Sci.*, 1952, **98**, 256–264.
38. HALSTEAD, H. A psychometric study of senility. *J. ment. Sci.*, 1943, **89**, 363–373.
39. HALSTEAD, H. Mental tests in senile dementia. *J. ment. Sci.*, 1944, **90**, 720–726.
40. HEBB, D. O. *The organization of behavior: a neuropsychological theory.* New York: John Wiley, 1949.
41. HOPKINS, BARBARA, & POST, F. The significance of abstract and concrete behaviour in elderly psychiatric patients and control subjects. *J. ment. Sci.*, 1955, **101**, 841–850.
42. HOPKINS, BARBARA, & ROTH, M. Psychological test performance in patients over sixty. II. Paraphrenics, arteriosclerotic psychosis and acute confusion. *J. ment. Sci.*, 1953, **99**, 451–463.
43. HULL, C. L. The formation and retention of associations among the insane. *Amer. J. Psychol.*, 1917, **28**, 419–435.
44. HUNT, J. MCV. Psychological experiments with disordered persons. *Psychol. Bull.*, 1936, **33**, 1–58.
45. INGLIS, J. An experimental study of learning and "memory function" in

elderly psychiatric patients. *J. ment. Sci.*, 1957, **103**, 796–803.

46. INGLIS, J., SHAPIRO, M. B., & POST, F. "Memory function" in psychiatric patients over 60. The role of memory in tests discriminating between "functional" and "organic" groups. *J. ment. Sci.*, 1956, **102**, 589–598.
47. JONES, H. E., & KAPLAN, O. J. Psychological aspects of mental disorders in later life. In O. J. Kaplan (Ed.), *Mental disorders in later life.* (2nd ed.) Stanford: Stanford Univer. Press, 1956.
48. KRAL, V. A., & DUROST, H. B. A comparative study of the amnestic syndrome in various organic conditions. *Amer. J. Psychiat.*, 1953, **110**, 41–47.
49. KRAWIECKI, J. A., COUPER, L., & WALTON, D. The efficacy of parentrovite in the treatment of a group of senile psychotics. *J. ment. Sci.*, 1957, **103**, 601–605.
50. LASSEN, N. A., MUNCK, O., & TOTTEY, E. R. Mental function and cerebral oxygen consumption in organic dementia. *Arch. neurol. Psychiat.*, 1957, **77**, 126–133.
51. LILJENCRANTS, J. Memory defects in organic psychoses. *Psychol. Monogr.*, 1922, **32**, 143.
52. LOVETT DOUST, J. W., SCHNEIDER, R. A., TALLAND, G. A., WALSH, M. A., & BARKER, G. B. Studies on the physiology of awareness. The correlation between intelligence and anoxemia in senile dementia. *J. ner. ment. Dis.*, 1953, **117**, 383–398.
53. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* (2nd ed.) New York: Longmans Green, 1952.
54. MOORE, T. V. The correlation between memory and perception in the presence of diffuse cortical degeneration. *Psychol. Monogr.*, 1919, **27**, 120.
55. ORME, J. E. Intellectual and Rorschach test performance of a group of senile dementia patients and of a group of elderly depressives. *J. ment. Sci.*, 1955, **101**, 863–870.
56. ORME, J. E. Non-verbal and verbal performance in normal old age, senile dementia and elderly depression. *J. Gerontol.*, 1957, **12**, 408–413.
57. PICHOT, P. Language disturbances in cerebral disease. *Arch. neurol. Psychiat.*, 1955, **74**, 92–95.
58. PINKERTON, P., & KELLY, J. An attempted correlation between clinical and psychometric findings in senile-arteriosclerotic dementia. *J. ment. Sci.*, 1952, **98**, 244–255.
59. RABIN, A. Psychometric trends in senility and psychoses of the senium. *J. gen. Psychol.*, 1945, **32**, 149–162.
60. RAVEN, J. C. *The Mill Hill Vocabulary Scale.* London: Lewis, 1948.
61. RAVEN, J. C. *Guide to using the Progressive Matrices.* London: Lewis, 1950.
62. ROBERTSON, J. P. S. Age, vocabulary, anxiety and brain damage as factors in verbal learning. *J. consult. Psychol.*, 1957, **21**, 179–182.
63. ROBERTSON, J. P. S., & WIBBERLEY, H. Dementia versus mental defect in middle-aged housewives. *J. consult. Psychol.*, 1952, **16**, 313–315.
64. ROTH, M., & HOPKINS, BARBARA. Psychological test performance in patients over 60. I. Senile psychosis and the affective disorders of old age. *J. ment. Sci.*, 1953, **99**, 439–450.
65. SHAKOW, D., DOLKART, M. B., & GOLDMAN, R. The memory function in psychoses of the aged. *Dis. nerv. System*, 1941, **2**, 43–48.
66. SHAPIRO, M. B., POST, F., LÖFVING, BARBRO, & INGLIS, J. "Memory function" in psychiatric patients over sixty, some methodological and diagnostic implications. *J. ment. Sci.*, 1956, **102**, 233–246.
67. SILVERMAN, A. J., BUSSE, E. W., BARNES, R. H., FROST, L. L., & THALER, MARGARET B. Studies on the processes of aging. 4. Physiologic influences on psychic functioning in elderly people. *Geriatrics*, 1953, **8**, 370–376.
68. SIMMINS, CONSTANCE. Studies in experimental psychiatry. IV. Deterioration of "g" in psychotic patients. *J. ment. Sci.*, 1933, **79**, 704–734.
69. SPEARMAN, C. *The abilities of man.* London: Macmillan, 1932.
70. STEPHENSON, W., PENTON, C., & KORENCHEVSKY, V. Some effects of vitamins B and C on senile patients. *Brit. med. J.*, 1941, **2**, 839–844.
71. THALER, MARGARET. Relationships among Wechsler, Weigl, Rorschach, EEG findings, and abstract-concrete behavior in a group of normal aged subjects. *J. Gerontol.*, 1956, **11**, 404–409.
72. TRUEBLOOD, C. K. The deterioration of language in senility. *Psychol. Bull.*, 1935, **32**, 735. (Abstract)
73. VERNON, P. E., & MCKINLEY, M. Effects of vitamin and hormone treatment on senile patients. *J. neurol. neurosurg. Psychiat.*, 1946, **9**, 87–92.

74. WECHSLER, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.

75. WECHSLER, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, **19**, 87–95.

76. WELLS, F. L., & MARTIN, HELEN A. A. A method of memory examination suitable for psychotic cases. *Amer. J. Psychiat.*, 1923, **3**, 243–257.

77. WILLIAMS, MOYRA. Studies of perception in senile dementia: Cue selection as a function of intelligence. *Brit. J. med. Psychol.*, 1956, **29**, 270–279.

78. WILLIAMS, MOYRA. Spatial disorientation in senile dementia. *J. ment. Sci.*, 1956, **102**, 291–299.

79. WYLIE, MARGARET. An experimental study of recognition and recall in abnormal mental cases. *Psychol. Monogr.*, 1930, **39**, 180.

80. YACORZYNSKI, G. Y. An evaluation of the postulates underlying the Babcock deterioration test. *Psychol. Rev.*, 1941, **48**, 261–267.

81. YATES, A. J. The use of vocabulary in the measurement of intellectual deterioration—a review. *J. ment. Sci.*, 1956, **102**, 409–440.

82. ZANGWILL, O. L. Clinical tests of memory impairment. *Proc. Roy. Soc. Med.*, 1943, **36**, 576–580.

83. ZILBOORG, G., & HENRY, G. W. *A history of medical psychology.* New York: Norton, 1941.

Received December 30, 1957.

PSYCHOLOGICAL BULLETIN
VOL. 55, NO. 4, 1958

RECENT STUDIES OF EYE MOVEMENTS IN READING[1]

MILES A. TINKER
University of Minnesota

Although there has been a decrease in the number of experiments undertaken during the last 11 years, the study of eye movements in reading continues to be an important technique for investigating the reading process. An earlier article by Tinker (**61**) plus his last review (**62**) to appear in this journal covers all the material in the field up to 1945. In the present review, articles appearing during the years from January 1945 to October 1957 are considered. A few reports not available to the reviewer have been included in the bibliography for the sake of completeness (**18, 21, 31, 32**). Bibliographies, critical evaluations, and summaries of parts of the field will be found in references (**2, 5, 7, 11, 17, 22, 27, 30, 34, 41, 44, 48, 53, 56, 61, 62, 67, 68, 71**). Certain materials on eye movements will not be reviewed or will be only mentioned briefly since they are not intimately related to oculomotor behavior in ordinary reading. These include reports on eye movements in viewing pictures or advertisements, certain of the studies on visual fixation, and eye movements as related to visual attention.

The studies to be reviewed here group themselves as follows: techniques of measurement, analysis of the reading process, training to improve eye movements, typography and eye movements, eye movements and fatigue, and summary statement.

TECHNIQUES OF MEASUREMENT

Most of the reports dealing with techniques of measurement have been concerned with instrumentation. Some authors describe modifications of earlier apparatus for recording eye movements; others set forth new principles of measurement. Brandt (**6**) reports the improvements in his bidimensional camera. It employs a corneal reflection technique with the camera lenses so arranged that the reflected beams of light come to a focus and are recorded as a tiny dot on the film. The 35mm film moves at a constant rate and stops intermittently so as to catch each new fixation. The camera photographs all eye movements in every direction on a single film. Analysis of the ocular pattern portrayed on the film yields the duration, location, and sequence of every fixation and every eye movement. The visual field to be examined by a subject can be any size up to a double-page spread of the *Saturday Evening Post*. This bidimensional camera provides an efficient, flexible, and reliable instrument for research in a variety of eye-movement studies including reading. Allen (**1**) also describes a relatively simple corneal reflection photographic technique for continuously recording vertical and horizontal eye movements on one 35mm film. The film runs continuously in the 45° meridian. A mask with small slits at right angles to the film is interposed between the camera lens and film. This produces a record of horizontal movements from one eye and vertical movements from the other eye. The apparatus may be used for a variety of research projects besides reading. However, as recognized by the author, the use of one eye as an indicator of vertical movements and the other eye for hori-

[1] The writer is grateful to the University of Minnesota Graduate School for a research grant to finance preparation of this material.

zontal movements is valid only when fusion, ductions, and phorias are normal and the visual field is close to a single plane. The Lifwynn eye-movement camera is described by Syz (**57**) and by Burrow and Syz (**14**). It is an elaborate and apparently expensive apparatus which photographs the corneal reflection on 35mm film. It is possible to expose frames from once to 60 or more times per second. Duration of fixation and sequence of movements are obtained from the records. Apparently this camera was designed primarily to study the influence of various factors on visual fixation. It is not readily applicable to reading research.

A number of reports describe electrical methods of recording eye movements: Brockhurst and Lion (**8**), Gemelli, Colombi, and Schupfer (**23**), Lundberg (**42**), Marg (**44**), Powsner and Lion (**49**), ten Doesschate and Lansberg (**58**). Carmichael and Dearborn (**17**) have described in an excellent manner representative methods developed to record eye movements of human subjects while reading. In the chapter on recording of eye movements (pp. 146–205) they describe in detail and evaluate electrical techniques. The interested reader is referred to this reference. An electrical method has distinct advantages in some experimental situations. This is particularly true when continuous records must be taken for long reading periods. In such situations, the less restraint placed on the reader the better. In contrast to the electrical methods the corneal reflection method imposes rather severe restraint upon the reader by head clamps and the necessity of elaborate apparatus in close contact with the subject. In general, recent trends have been to use an electrical rather than a corneal reflection method. Nevertheless, for certain purposes in studying eye movements a refined photographic technique (corneal reflection) is probably superior to electrical recording (see Tinker [**62**]). Hartridge and Thomson (**30**) consider that photographing the reflection from a scleral mirror is superior to photographing the corneal reflection or the edge of the iris in studying eye movements. However, any method, including this one, that involves an attachment to the eyeball has distinct disadvantages.

Analysis of the Reading Process

Unit of perception in reading. In a study by Muñoz, Odoriz, and Tavazza (**46**), the eye movements of children were recorded by means of a Grass Electro-Encephalograph while reading Spanish. It was found that the unit of recognition in reading was a word or group or words. And when the reading rate was increased through specific training, there was a corresponding increase in the amount recognized at each fixation of the eyes. More difficult material was read with more fixations of the eyes. It was concluded that the natural form of reading is not by spelling or syllabizing but on the basis of whole groups of words. That is, word forms or configurations constituted the units of perception in reading.

As part of a larger study, Gray (**28**) compared the basic reading process of mature readers in 14 different languages by an analysis of eye-movement records. The languages were Arabic, Burmese, Chinese, English, French, Hebrew, Hindi, Japanese, Korean, Navaho, Spanish, Thai, Urdu, and Yoruba (native Nigerian). There were two to seven subjects in each language group. The eye movements were photographed with the University of Chicago (corneal reflection) camera. The content of the passages read was the same for all

subjects except for the Navaho Indians. Records were obtained for both oral and silent reading. The data indicate that the general nature of the reading act is essentially the same among all mature readers. As the mature reader seeks the meaning of the passage, he follows along the lines with an alternation of short eye movements and pauses. At each fixation pause he recognizes words as wholes, usually two or three at a time, and by means of their configuration and striking characteristics. Occasionally regressions occur. Although there will be some variation due to form and structure of a language, the teaching of the basic skills in all of them can be similar to some degree. This is an important contribution to the study of the nature of the reading process. The conclusions are supported by other studies of reading Chinese and Japanese cited by Gray (**28**). A recent study of eye movements in reading German and English by Waterman (**70**) is also in agreement with Gray. The former found no discernible variation between the eye-movement reading patterns of literate native speakers of these languages. Also there is no apparent change in reading habits (eye movements) when a native speaker of one language learns to read well in another language.

Accuracy in visual fixation and movement. Certain studies of visual fixation, speed of eye movements, and vision during eye movement have some bearing on the study of eye movements in reading. Riggs, Ratliff, Cornsweet, and Cornsweet (**51**) investigated the role of the disappearance of a steadily fixated visual test object when the image is maintained in a constant location on the retina during normal involuntary movements of the eye, and when normal fixation movements are effectually doubled. Results indicate that the normal eye movements during fixation serve to overcome loss of vision which would result from prolonged, uniform stimulation. This would probably have a bearing in special reading situations which require prolonged fixation as studying an intricate formula, especially when the eyes are fatigued. Lord and Wright (**40**) studied the eye movements occurring during fixation. They discovered rapid flicks lasting .02 to .03 sec. with amplitudes of 2 to 14 minutes arc.

Saccadic eye movements, such as those employed in reading, appear to be remarkably constant in speed. Brockhurst and Lion (**8**) found no decrement or acceleration during 90° sweeps at rates of 120 to 150 times per minute, although near the end of the test the moves were less rhythmical, or the eyes did not sweep the full distance. Kleitman and Schreider (**35**) did find some diurnal variation for saccadic sweeps in subjects kept awake for 24 hours. Performance was poorest in the early morning. Lateral sweeps templeward were faster than nasalward. This latter finding had been reported by Miles in 1924 (see Tinker [**61**]). Binocular saccadic eye movements are ordinarily well coordinated. This has been shown by Lord (**39**) and Brandt (**6**). It is well known that there is no clear vision during saccadic eye movements. Bell and Weir (**4**) suggest that this inhibition takes place in the cortex. This agrees with Holt (1903) but not with Dodge (1900) (see Tinker, [**61**]).

Tinker (**63**) has coordinated the results of the Minnesota studies on the time relations for eye-movement measures. The results follow: (*a*) Characteristics of saccadic eye movements are essentially the same whether they are interfixation movements in reading or other excursions

in the visual field. (*b*) There are significant individual differences in speed of saccadic eye movements. (*c*) Maximum velocity during eye movements of large amplitude is greater than that for small amplitudes. (*d*) Interfixation eye movements in reading take 10 to 23 msec. and the return sweep to the next line, 40 to 54 msec. (*e*) For most reading situations, eye movements take 6 to 8% of reading time while the rest (92 to 94%) is devoted to pauses, the periods of clear vision.

Pause duration in reading. Various aspects of pause duration patterns in reading are considered by Tinker (**64**). The first problem was to determine minimum pause duration for seeing clearly during binocular vision. It was found that it took on the average 172 msec. to fixate a dot at the end of a saccadic move, and 157 msec. to fixate similarly and identify a letter. These durations are greater than necessary where no saccadic move is involved. A tachistoscopic exposure of 100 msec. is adequate for a well cleared-up perception (see Tinker [**61**]). In reading, pause duration is even longer and depends upon the content of the material read. Thus, for easy prose it is about 220 msec., for scientific prose about 236 msec., and for reading objective test items about 270 to 324 msec. on the average. In general, pause duration fluctuates around 250 msec. for adults in ordinary reading. The reasons for longer pause duration in reading than for cleared-up vision during fixation as described above are: Oculomotor adjustments of the eyes constitute one set of factors. First, we find that the mean reaction time of the eyes to eccentric stimulation is about 173 msec. (**64**). Note that this is about the same as the fixation pause on a dot at the end of a saccadic move (above). Secondly, the eyes converge during saccadic interfixation movements and diverge during the fixation which follows the movement. The eye must complete this divergence to achieve clear vision at the fixation and this takes time. Another set of factors involves the comprehension process. In addition to seeing clearly during a fixation pause, the reader must comprehend the ideas and relationships involved. Actually, therefore, pause duration includes perception time plus thinking time. For instance, in reading without attention to meaning, pause durations are brief and constant while for reading algebraic problems they are long and variable.

In reading simple material, pause duration is relatively constant but in more complex materials it can be highly variable. Tinker (**64**) shows that the coefficient of variability ranges from 10.4 to 25.8. He also shows that reliability of pause duration is high, i.e., .82 to .85.

Pause duration also varies with changes in the typographical arrangement. This will be discussed below.

In the same report, Tinker showed that pause duration alone has poor validity as a measure of reading proficiency. However, when combined with fixation frequency to produce perception time, the latter is fairly valid as a measure of speed of reading proficiency.

Reaction time of the eye. As noted above, reaction time of the eye to eccentric stimulation is involved in reading movements. Tinker (**64**) reports an average reaction time of 173 msec. As early as 1908 Diefendorf and Dodge, cited by Walton (**69**), reported reaction saccadic movements to a stimulus appearing in indirect vision to be 195 msec. on the average with a range of 120 to 235 msec. Westheimer (**71**) reports reaction times of the eye to eccentric

stimulation by neon bulbs to be 120 to 180 msec. Gerathewohl and Strughold (**24**) report a reaction time of 250 msec. Walton (**69**) devised a clever experiment to duplicate eye movements and fixations as utilized in a reading task while eliminating the factor of comprehension. This provided a means of determining the reaction time per fixation. Lines of five three-letter words, with one-half inch between words, were presented to 10 subjects before an eye-movement (Ophthalm-O-Graph). Records were obtained as the reader quickly fixated each of the successive words. Mean reaction time of the eye varied from 170 to 309 msec. with a group average of 219 msec. From these data on reaction time, Walton computed the maximum reading rates possible in terms of limiting physiological and anatomical factors. With a reaction time of 200 msec. and 3 or 2 fixations on a ten-word line, the rate would be 853 or 1250 words per minute (W.P.M.). And with a reaction time of 167 msec. and 3 or 2 fixations per line, the rate would be 982 or 1451 W.P.M. He concludes that for rates over 1451 W.P.M., the reader is skimming. These figures appear to be overoptimistic since no allowance is made for comprehension time which must enter into pause duration time in ordinary reading. No published reports cite average reading pauses as low as 200 msec. to say nothing of 167 msec. Thus Dixon's Case 13 (**19**), who had the reputation of reading whole lines or paragraphs at a glance, only read about 500 W.P.M. and used three or more fixations per line. The reviewer suggests that any rate of over about 800 W.P.M. can only mean that the reader is skimming rather than reading all the material. Even with easy material, 500 W.P.M. is very fast reading. Only the exceptional individual could achieve around 800 W.P.M. Any claims by advertised reading improvement programs that one can learn to *read* at 1500 to 3000 or more W.P.M. can only be false.

Rhythm reading. When readers tend to employ approximately the same eye-movement pattern from line to line they are said to employ a rhythmical pattern of eye movements. Dixon (**19**) found evidence of rhythmical reading in only a few of the records of his good readers. See Tinker (**62**, p. 98) for an evaluation of this concept of rhythm reading. According to recent studies the view that rhythmical eye-movement patterns are desirable for effective reading becomes meaningless. Training designed to produce rhythmic patterns of eye movements in order to improve reading, therefore, is based upon an invalid assumption.

Vertical vs. horizontal reading. Studies of reading Chinese and Japanese in vertical and horizontal alignment have been considered in previous reviews (**61**, **62**). In Tinker's study (**65**) of reading English material in vertical and horizontal arrangements, 10 college subjects were given practice during six weeks in reading material in a vertical arrangement (42 readings of 300-word selections). Prior to this practice, eye movements were photographed while reading comparable selections in both the horizontal and vertical arrangements. The same was done after the six weeks of practice. At the initial testing, the vertical reading was 50% slower than the horizontal, but after practice it was only 21.8% slower. On the initial test, vertical reading required fewer fixations, fewer regressions, and longer pause durations. Practice produced marked improvement (reduction) in fixations and regressions. Apparently long established reading habits produced the

superiority of the horizontal reading. With longer practice, however, it is likely that the vertical reading would equal or excel the horizontal. It is noteworthy that more words were read per fixation and fewer regressions occurred in the vertical arrangement even before practice. And as pointed out by Shen (see Tinker, **61**), vertical eye movements may be better adapted than horizontal movements to the short interfixation movements used in reading.

Oculomotor efficiency and reading. Gilbert (**25**) has attempted to study the relation of growth in simple oculomotor control in relation to growth of eye movement patterns in reading. Growth in oculomotor control was recorded by photographing eye movements while reading series of digits and growth in eye-movement patterns while reading simple prose. Pupils in Grades I through IX were used as subjects. The number in each grade ranged from 18 to 65 for prose reading; from 29 to 71 for digit reading. A group of 42 college students was also used. Records were obtained for reading 100 words and 100 digits.

It is claimed in the discussion that skill in directing the eye movements in reading digits (called simple motor activity) is substantially related to oculomotor performance in reading connected (simple) material. Details of the findings and conclusions need not be reported here because it seems to the reviewer that the author has employed an invalid criterion of coordinated motor activity of the eyes. Digit *reading* is classified as simple motor activity of the eyes. But reading digits is reading. Pupils were shown how to read the digits prior to photographing their eye movements. After all, digits are symbols for words. Except for space required for printing, *9* is the same as *nine*. The series of digits were read, not just fixated. The prose samples must have been extremely easy reading for all except the first and perhaps the second grade pupils. Eye movements in reading such material, therefore, should be little influenced by comprehension factors. So there is little surprise that oculomotor patterns for reading the digit series and the easy prose turned out to be similar. Undoubtedly they are similar reading situations.

A number of other specific conclusions are at fault. Following are a few: (*a*) Eye-movement patterns do not reflect efficiency of central processes of comprehension. There is plenty of contrary evidence in the literature. If the author had presented both easy and very difficult materials to his readers he would have obtained evidence that difficulties of comprehension are reflected in eye movements. (*b*) Eye-movement records do not predict reading test performance. They would have if taken while reading the test material (see Tinker [**62**, p. 95]). (*c*) Poor oculomotor coordination is a handicap in learning to read well. Since no satisfactory measure of this coordination was used, the conclusion does not follow (see Tinker [**62**, pp. 110–114]).

The author has made a good contribution to development of eye-movements with age. This will be considered below.

Eye-movement changes with age. In the study discussed just above, Gilbert (**25**) obtained eye-movement measures for fixation frequency, regressions, and pause duration during reading for Grades I through IX and for college students. The same prose selection (easy) was used at all levels except college. Proficiency in all measures improved throughout the

educational levels although most of the gains were achieved by the fifth grade. For good readers the most rapid growth is during the first four grades. Poorer readers continue to gain throughout most of the grades. Oculomotor behavior of college students was only slightly more mature than for ninth-grade pupils. Growth of eye-movement patterns in reading series of digits showed similar patterns. Much the same growth patterns in eye movements were discovered when the same pupils were photographed yearly for three successive years in Grades II through IV and IV through VI. Thus, records from samples at various grade levels apparently yield valid measures of oculomotor growth.

Ballantine (**3**) photographed eye movements for readers in Grades II, IV, VI, VIII, X, and XII to obtain growth curves for the various eye movement measures (fixations, regressions per em, regressions per line). All subjects (20 in each grade) read an easy selection (second-grade difficulty) and a selection appropriate to their own grade in difficulty. The experiment was carefully controlled in all respects. Fixation and regression frequency improved rapidly from Grades II to IV, slower from IV to VIII, and only slightly or not at all from VIII to XII. Refixations per line decreased steadily to Grade VIII. These changes tended to be significant to Grade VIII but not at higher levels. There were no important differences in the measures between the easy selection and the at-grade selections. The author's conclusion that this indicates that difficulty of material does not affect eye movements is not valid. The difficult selections were *at grade* for the pupils. The results only signify that the eye movements are approximately the same for reading material at grade and material that is easier than at grade.

Morse (**45**) investigated the eye movements of fifth- and seventh-grade pupils when they read material at grade (in difficulty), at two grades below, and at two grades above their school placement. There were 54 readers in each grade. The reading selections were carefully equated to the proper level. The eye-movement patterns of the seventh-grade pupils were more efficient than those of fifth-grade on both seventh- and fifth-grade material. And seventh-grade pupils were more efficient in reading fifth-grade material than fifth graders reading third-grade material. Eye-movement patterns for both fifth- and seventh-grade pupils did not change in any predictable way with increase in difficulty of material by two grades. This finding is indeed unfortunate. It indicates that the school children used as subjects had not been successfully taught to vary their pace according to the difficulty of the material, i.e., they were not flexible in adjusting reading procedure to difficulty of material. It is unlikely that the purpose for the reading was so rigidly set that it would have prevented flexibility in adjusting to difficulty of material. When eye movements do not vary with difficulty of reading matter, pupils are immature readers.

As part of an extensive investigation, Dunn (**20**) compared the eye movements of retarded boys with those of normal boys of the same mental age. There were no significant differences between the two groups in fixation frequency, regressions, and rate of reading. The comprehension of retarded boys was poorer than for the normal boys. The former apparently merely looked at

words without understanding the content.

Effect of type of material and set. In Dixon's study (**19**), a group of professors and a group of graduate students in physics, in history, and in education read passages in their own and in each of the other two areas. The passages were equated for difficulty by standard formulas. There were 16 subjects in each subgroup. Eye movements were photographed while the passages were read. In considering the results, one should keep in mind that the subjects were highly selected experts. The conclusions cannot be applied to the reading of school children. Professors, and to a lesser degree the graduate students, tended to read material in their own field with more efficient oculomotor patterns, i.e., familiarity of material is a factor in determining reading performance. Different types of material read with the same directions (set) do not automatically elicit different patterns of eye movements *when the passages are equally difficult.* These findings are important. The above findings do not mean, however, that reading for different purposes or reading materials with wide variations in difficulty would not produce variation in oculomotor behavior. Care, therefore, must be exercised in interpreting the findings of this study. Ordinarily the material that is read in physics or mathematics at any grade level is not of the same difficulty as that read in the social studies or literature in the same grade. Furthermore, the purposes for which the reading is done tends to differ from one subject matter to another.

Ledbetter (**37**) investigated the variation in eye-movement patterns of 60 eleventh-grade pupils while reading selections in English, mathematics, natural science, and social science. An attempt was made to equate the difficulty of the passages read in terms of number of words, vocabulary difficulty, sentence length, and grammatical structure. No standardized formula was used for this. Significant differences in oculomotor patterns were found for reading in the various subject matter fields. The eye movements in reading a poem and mathematics were more complex than in the other areas. The author admits that the method of equating materials for difficulty may not have been valid. Other earlier studies such as those of Seibert and of Stone (see Tinker [**62**]) found differences in eye-movement measures while reading different subject matter. Examination of all the literature suggests that variation in eye movements occurring with variation in subject matter is due largely to differences in difficulty of the material or to changes in the purpose for which the reading is done rather than to differences in subject matter as such. For instance, in Dixon's study cited above, the more efficient reading of materials that are familiar to the subject may only mean that the familiar material was easier even though equated to other materials for difficulty by a formula.

In an unpublished study by G. R. Klare, E. H. Shuford, and W. H. Nichols[2] the eye movements of 30 college students were photographed while reading technical material differing in style difficulty first with a weaker set to learn and also with a stronger set to learn. The easier style was read with significantly fewer fixations and regressions. The stronger set to learn resulted in somewhat

[2] Communicated to the writer in mimeograph form by G. R. Klare.

more fixations and regressions. Thus style difficulty is inversely related to reading efficiency as measured by eye movements.

Flexibility in adjusting eye movements. It is generally recognized that the mature reader is the versatile reader. He will change his pace (reflected in eye movements) to fit the purpose of reading and the nature and the difficulty of the material. He will read rapidly when that is appropriate. In certain other situations he will employ slow, analytical reading. Many readers have not achieved this versatility (see the Morse study reviewed above). Laycock (**36**) investigated what happens to oculomotor behavior of good readers when they try to read simple material at a very rapid rate. A sample of 72 was drawn from 391 college readers: 37 flexible and 35 inflexible. The flexible readers were able to read quickly when directed to and maintain adequate comprehension; the inflexible ones could not change their pace even when directed to do so. Eye movements of these two groups were photographed when reading at normal rate and when directed to read quickly. Analysis of the eye-movement measures showed that the flexible in comparison to the inflexible group had significantly fewer fixations and shorter pause duration when directed to read quickly. There was no significant change in regressions. The inflexible group did make some improvement but far less than that made by the flexible readers. So readers tend to be more or less flexible and inflexible rather than strictly flexible and inflexible. It is suggested that emphasis be placed upon helping the less flexible readers in reader improvement programs since they tend to do all their reading at about the same rate. Lack of flexibility in reading at the college level constitutes a serious handicap.

Individual differences. Wide individual differences are always found in reading. These are readily detected in eye-movement records. Special emphasis is placed upon these variations by such authors as Morse (**45**), Dixon (**19**), Ballantine (**3**), and Gilbert (**25**).

Eye movements in special reading situations. Brandt (**6**) cites data on eye movements while reading in several special reading situations: (*a*) *Geometry.* In comparing good and poor geometry students at the end of two semesters' study, the poor students made about three times as many fixations as the good students per correct answer to problems. The many random eye movements of the inferior students indicated lack of systematic attack on problems. (*b*) *Algebra.* Here also the inferior students made more fixations and exhibited random moves for re-examination of the problems. (*c*) *Arithmetic.* In attempting to find an error in a long division problem, superior students use more systematic patterns of eye movements. (*d*) *Spelling.* In attempting to identify correctly spelled words in a multiple-choice test, an inefficient (and low ability) pupil employed inadequate sequences of moves. Even the good student used many random and haphazard eye movements. (*e*) *Geography.* In reading a map, the efficient pupil employs relatively few systematic fixations and eye movements. (*f*) *Intelligence tests.* Low ability pupils make more fixations and more ineffective moves than high ability pupils. In general, in these specific learning situations, the efficient pupil tends to employ effective oculomotor patterns while the inefficient pupil uses an excessive number of fixations and a hap-

hazard pattern of movements. These conclusions must be tentative since relatively few subjects were measured by Brandt.

Lofquist (**38**) studied the eye-movement patterns in reading clerical test items which consist of pairs of names and pairs of numbers to be compared. The eye movements of 40 university students were photographed while they read 16 lines of prose, 20 name and 20 number items of the Minnesota Vocational Test for Clerical Workers. Response to the test required detection of likeness or difference of the two parts in an item. Reading the test items required a more analytical procedure than prose, i.e., more fixations and more regressions. The number items were read with more fixations and regressions and longer perception time than the name items. A similar trend was found for long (more complex and difficult) vs. short items. The results of this study furnish confirmation of earlier investigations (**61**, **62**) which have demonstrated the adaptability of eye movements to central perceptual processes. It is now well established that oculomotor reactions are exceedingly flexible and quickly reflect any variation in the central processes of perception, judgment, comprehension, etc. In other words, it appears that eye-movement patterns merely reflect ease or difficulty of reading, efficient or poor reading performance, and degree of comprehension, rather than cause good or poor reading. Certain writers, on the basis of incomplete evidence, wrongly infer that eye-movement patterns are stable and unmodifiable (**45**) or are limited and fixed by some underlying motor ability (**25**). Versatility in adjusting reading habits (including eye movements) to variation in purposes and materials is one "hallmark" of maturity in reading.

Training to Improve Eye Movements

Various kinds of training have been employed in attempts to improve eye movements in reading. It is assumed by some authors that if a person's oculomotor patterns are developed to be similar to those which characterize efficient reading, his reading proficiency would improve. The earlier studies in this area have been evaluated by Tinker (**62**). The more recent investigations will be briefly noted. Glock (**26**) studied the effect upon eye movements of three methods of training: (*a*) using the Harvard films which expose phrases in succession and thus train eye movements; (*b*) using a new film which exposed two successive lines simultaneously; and (*c*) reading printed material while motivated to read fast and comprehend. Four weeks training was given to six sections of college students. Students made significant improvement in eye movements (fixations, regressions, pause duration) under all three methods of training. But there were no significant differences between the results of the three methods, i.e., the technique that paced the eyes was no more effective than the others in modifying eye movements.

In a carefully controlled experiment, Manolakes (**43**) checked the influence of omitting tachistoscopic training on changes in eye movements. Experimental and control groups received training on a reading rate controller plus vocabulary and comprehension training. The control group also had tachistoscopic training; the experimental group did not but received somewhat broader training in vocabulary and comprehension. End tests revealed no significant differences between groups in reduction of fixations, regressions, and pause duration. Therefore, the read-

ers were not penalized by omission of tachistoscopic training. This substantiates previous findings that tachistoscopic training tends to be of doubtful value in the reading improvement program.

In Tillson's experiment (**59**), college students in a semester reading course were trained with a reading accelerator, the Harvard films, and timed reading. Eye-movement records taken near the beginning and near the end of the course revealed a significant reduction in fixation frequency and regressions. There were also significant gains in speed and comprehension. It was concluded that reading proficiency was improved by changing reading (oculomotor) patterns. The conclusion should be that the improved reading was reflected in a changed oculomotor pattern. Like so many authors, Tillson has assumed wrongly that proficient eye movements produce efficient reading rather than vice versa.

Westover (**72**) investigated the comparative effectiveness of three methods of improving the reading performance of college freshmen: (*a*) Group I had two fifty-minute practice periods per week for five weeks in reading and taking tests on study type of reading exercises; (*b*) Group II had had the same practice on the same material by means of a device for controlling eye movements; (*c*) Group III attended college but received no special instruction in reading. All groups made significant gains in speed and comprehension but Groups I and II made greater gains. There was, however, no difference in the gains for Groups I and II. Therefore, the mechanical control of eye movements did not show significantly better results than the same reading exercises used alone.

The present evaluation of eye-movement training or training to provide improved eye movements in reading would be the same as that given by Tinker (**62**, p. 112) in 1946. The improvement obtained by such training, with or without elaborate apparatus, is no greater than that resulting from well-motivated reading practice alone. See Tinker (**62**) for details.

Typography and Eye Movements

Although Burt, Cooper, and Martin (**15**) considered eye movements in studying the effect of typographical variations on reading, no important results are cited. Hackman and Tinker (**29**) made an analysis of oculomotor patterns employed during the reading of material printed in a variety of combinations of colored ink and colored paper. A latin square experimental design and 49 readers were used. Each of the eye-movement measures (perception time, fixation frequency, pause duration, and regression frequency) showed variations from one color combination to another in such a way as to indicate print of good and of poor legibility. Black (ink) on yellow (paper), red on white, green on red, and black on white provided best legibility. Black on purple, pale orange on white, and red on green provided worst legibility. The results indicate that brightness contrast between ink and paper rather than color variation as such determines legibility of the print. In using combinations of colored ink and paper, therefore, maximum legibility is achieved by using a printing arrangement with a maximum brightness contrast between print and background. It might be mentioned that these findings are in harmony with a large body of data derived from experiments on brightness contrast in relation to visibility of print and speed of reading.

In the reading situation the amount of material perceived at each fixational pause is known as the perceptual or fixation span. Paterson and Tinker (**47**) studied the effect of typographical variations upon the perceptual span in reading. The following nonoptimal typography reduced significantly the perceptual span: Old English type face, 6 point and 14 point type, 9 and 43 pica line widths, a low brightness contrast between print and paper, and combinations of the above. In general, optimal typography favors a large perceptual span and nonoptimal typography reduces significantly the span. It should be noted that other factors such as comprehension requirements may affect the perceptual span more than typographical changes.

An extensive study of the effect of typographical variations upon eye movements in reading was completed by Tinker and Paterson (**66**). Eye movements were photographed while reading optimal and nonoptimal printing arrangements. The typographical variations studied included line width, size of type, type faces, type form, white vs. black print, brightness contrast between print and paper, and combinations of the above. The oculomotor measures were fixation frequency, words per fixation, regression frequency, pause duration, and perception time. The oculomotor patterns varied from one typographical arrangement to another. Examination of these patterns revealed the nature of the perceptual difficulties involved in each nonoptimal printing arrangement. For example, in contrast with lower case print, all capital printing was read with more fixations, a decrease in pause duration, an increase in perception time, and no change in regression frequency. The characteristic word forms which facilitate rapid reading in lower case printing are absent in all capitals. For details of the other comparisons, see the original report (**66**).

Eye Movements and Fatigue

The results in various early experiments (see Tinker [**62**]) suggest that efficiency of oculomotor behavior may be affected by conditions which are intended to produce fatigue. Hoffman (**33**) had 30 college subjects read continuously for four hours while eye movements were recorded as electro-oculargrams. Nine five-minute samples of the reading were recorded: at the beginning and during the last five minutes of each 30 minutes of reading during the four hours. The number of fixations (for five minutes) and the number of lines read decreased significantly after the first half hour of reading. However, length of reading period had relatively little influence on number of fixations per line, i.e., no significant increase until the very end of the four hours where there was an increased 0.430 of a fixation per line on the average. Apparently the author has misinterpreted his data. A decrease in fixations ordinarily means more efficient reading and an increase means less efficient reading. Since there was a decrease in lines read at successive half hours and since there were just about the same number of fixations per line at all times, there would necessarily be less fixations in five minutes of reading. So the decrement in total fixations per five minutes is an artifact. There were no significant changes in eye fixations per words read or per line except at the very last period measured in comparison to the first five minutes of reading. Furthermore, the decrement in number of lines read on successive

measurements may only mean a lessening of motivation as the author recognizes. Similarly regressions per line show no significant change until at the end of four hours of reading when there was an increase of 0.175 per line on the average. Considering the influence of long periods of reading on eye movements, the only significant finding in this study is that fixations per line and regressions per line increased significantly after four hours of reading. This suggests that fatigue was beginning to operate by the end of four hours of reading, but experimental confirmation is needed.

Carmichael and Dearborn (**17**) set themselves the task of determining how long a normal human subject can continue to read before there are significant changes in his reading behavior. Twenty subjects read an interesting historical novel for six hours continuously and for another six hours an economic treatise. Another 20 subjects read the same materials reproduced on microfilm (and projected) for similar periods. Short comprehension tests were interspersed during the reading to maintain motivation and a long test was given at the end of the six hours. Eye movements were recorded electrically in the form of an electro-oculogram. Five-minute records were obtained at the beginning and during the last five minutes of each half hour. Oculomotor measures considered were fixations per five-minute period, fixations per line, fixation sigma score, regressions per five-minute period, and regressions per line. The results, considering all oculomotor measures, indicate that subjects read as well at the end of six hours as at the beginning, i.e., the oculomotor patterns did not change significantly. This was true for reading both the books in regular print and in microfilm reproduction. It is suggested that the task of reading for six hours is such that for normal subjects a new "steady rate" of some sort may be established during continuous work by the visual receptor-neuromuscular mechanism. Except for a few reports of mild discomfort, it was not found that reading for the six hours was done at any "cost" to the organism. The practical educational inference that seems justified is that there is no basis for the belief that requiring long periods of reading by high school and college students may be injurious to the visual mechanisms of such students if their eyes are in fair condition to start with. One should keep in mind that the reading in this experiment was done under optimal or near optimal conditions of print and lighting. Furthermore, the authors do not deny that significant deterioration might occur under more stressful conditions. Brozek (**9**), in commenting on the findings, questions the author's interpretations and at the same time states that in a similar study Hoffman (**33**) obtained definite symptoms of deterioration within *four* hours. Actually, as stated above, Hoffman only found a suggestion of deterioriation at the very end of the four hours. Also the reader should remember that Carmichael and Dearborn limit fatigue as they use it to a descriptive term. Actually, in terms of this definition, therefore, and in view of the purpose of the experiment (see above), as well as in terms of practical implications (see above), Carmichael and Dearborn have made a very important contribution.

In considering quantitative criteria of oculomotor performance in relation to fatigue, Brozek (**10**) notes that in the course of severe stress of making saccadic eye moves as fast as possible for four minutes through an

arc of about 14° there was significant deterioration in fixation time, velocity of movements, and extent of corrective movements. After two hours of severe visual work, there was similar deterioration. Brozek, Simonson, and Keys (**13**) cite similar trends.

Simonson and Brozek (**54**) photographed saccadic eye movements before and after two hours of severe visual discrimination under 5, 100, and 300 foot-candles of illumination. There was deterioration in all oculomotor measures but only significantly so for part of the measures. In another study Simonson and Brozek (**55**) were concerned with the effect of spectral quality of light on visual performance and fatigue. One aspect dealt with oculomotor behavior before and at the end of two hours of severe visual work under five foot-candles of light. Although there was deterioration for most oculomotor measures, only two were significantly different. Velocity and average deviation of eye movements from the target were significantly greater for work under illumination from natural white lamps than for work under illumination from ordinary inside frosted lamps. The third type of lamp was the Verd-A-Ray. Later Brozek and Simonson (**12**) present a more detailed analysis of data comparing these three sources of illumination. No significant differences in oculomotor behavior were found after the two hours of severe visual work.

Eye movements of normal and exophoric (eyes tend to turn out from desired position) readers were photographed by Salatini (**52**). At the beginning of lines of print, i.e., at the end of the return sweep, the exophoric group made somewhat greater divergent movements. It is suggested that this might well account for the greater fatigue from reading of this group.

Evaluation of data on oculomotor behavior as a measure of fatigue (irrespective of how fatigue is defined) would seem to indicate the following:

1. In the ordinary reading situation with motivated readers at high school and college level, working under optimal conditions of print and light, oculomotor behavior does not deteriorate while reading continuously as long as six hours. There may be slight signs of fatigue after four hours of reading where motivation is not controlled.

2. Various aspects of saccadic eye movements (rate, accuracy of fixation at end of excursion, etc.) tend to show deterioration when under stress such as after two hours of strenuous visual inspection.

Summary Statement

Studies of eye movements in reading are diminishing in number during recent years. There are approximately 40% fewer studies during the past 10 years than in the previous decade. Recent studies reveal a lively interest in devising new or modified techniques for recording eye movements. There is a general tendency to move in the direction of some type of electrical recording to obtain electro-oculograms or something comparable. Photographing the corneal reflection is employed much less than formerly because it is less flexible than many of the newer techniques. A large amount of recent work has been concerned with the study of visual fixation, speed of eye movements, reaction time of the eye, oculomotor efficiency, and vision during eye movements.

Several experiments deal with the nature of the reading process, maturation of oculomotor patterns, and the influence of subject matter on eye movements. An appreciable number of investigations deal with the effects

of visual fatigue on eye movements and with the influence of typographical variations on oculomotor patterns. There are a few studies of eye movements in special reading situations. It appears that some writers still adhere to the mistaken notion that training eye movements as such is an effective way to improve reading. Numerous nonexperimental articles are concerned with general discussions and evaluations of eye movements in reading. It would appear from a survey of the literature that the study of eye movements in reading is to some degree reaching the stage of diminishing returns. Relatively few of the recent experiments deal with fundamental problems. Too many investigators appear to be unfamiliar with the literature in the field. In several instances findings are reported as new when in fact they are the same as discovered and reported earlier. The future study of eye movements in reading does not appear to be too promising. The eye-movement approach to the study of the reading process has probably made its major contributions. What we now need is less activity by dilettantes who are inadequately prepared to see the fundamental problems and unable to design suitable experiments in the field. Undoubtedly there will be a few researchers thoroughly grounded in the field who will sense basic problems and organize sound experimental procedures of investigation. Such individuals will be the ones to contribute worthwhile new information concerning eye movements in reading.

REFERENCES

1. ALLEN, M. J. A simple photographic method for continuously recording vertical and horizontal eye movements. *Amer. J. Optom.*, 1955, **32**, 88–93.
2. ANDERSON, I. H., & MORSE, W. C. The place of instrumentation in the reading program I. Evaluation of the Ophthalm-O-Graph. *J. exper. Educ.*, 1946, **14**, 256–262.
3. BALLANTINE, F. A. Age changes in measures of eye-movements in silent reading. In *Studies in the psychology of reading*. University of Michigan Monographs in Education, No. 4. Ann Arbor: Univer. of Michigan Press, 1951, 65–111.
4. BELL, G. H., & WEIR, J. B. DE V. Vision during glancing movements of the eyes. *Trans. Ophthal. Soc. U.K.*, 1947, **67**, 221–228.
5. BRANDT, H. F. Ocular photography as a scientific approach to the study of the psychological aspects of seeing. *Illum. Engng*, N. Y., 1944, **39**, 279–289.
6. BRANDT, H. F. *The psychology of seeing*. New York: Philosophical Library, 1945.
7. BRATBAK, J. Øyebevegelesene ved lesning. [Movements of the eyes in reading.] *Norsk ped. Tidskr.*, 1943, **27**, 59–163.
8. BROCKHURST, R. J., & LION, K. S. Analysis of ocular movements by means of an electrical method. *A.M.A. Arch. Ophthal.*, 1951, **46**, 311–314.
9. BROZEK, J. Visual fatigue: a critical comment. *Amer. J. Psychol.*, 1948, **61**, 420–424.
10. BROZEK, J. Quantitative criteria of oculomotor performance and fatigue. *J. appl. Physiol.*, 1949, **2**, 247–260.
11. BROZEK, J. Quantitative analysis of voluntary eye movements. In Gerard, R. W., *Methods in Medical Research*, Vol. 3. Chicago Yearbook Publisher, 1950, 199–207.
12. BROZEK, J., & SIMONSON, E. Visual performance and fatigue under conditions of varied illumination. *Amer. J. Ophthal.*, 1952, **35**, 33–46.
13. BROZEK, J., SIMONSON, E., & KEYS, A. Changes in performance and in ocular functions resulting from strenuous visual inspection. *Amer. J. Psychol.*, 1950, **63**, 51–66.
14. BURROW, T., & SYZ, H. Studies with the Lifwynn eye-movement camera. *J. Biol. Photogr. Ass.*, 1949, **17**, 155–170.
15. BURT, C., COOPER, W. F., & MARTIN, J. L. A psychological study of typography. *Brit. J. Stat. Psychol.*, 1955, **8**(1), 29–57.
16. CARMICHAEL, L. Reading and visual

work: a contribution to the technique of experimentation on human fatigue. *Trans. N. Y. Acad. Sci.*, 1951, **14**, 94–96.

17. CARMICHAEL, L., & DEARBORN, W. F. *Reading and visual fatigue.* Boston: Houghton Mifflin, 1947.
18. CORNSWEET, T. N. A determination of the stimuli for involuntary drifts and saccadic eye movements. *Dissertation Abstr.*, 1955, **15**, 1446.
19. DIXON, W. R. Studies of the eye-movements in reading of university professors and graduate students. In *Studies in the psychology of reading.* University of Michigan Monographs in Education, No. 4. Ann Arbor: Univer. of Michigan Press, 1951, 113–178.
20. DUNN, L. M. A comparison of the reading processes of mentally retarded and normal boys of the same mental age. In *Studies of reading and arithmetic in mentally retarded boys.* Child Development Publications of the Society for Research in Child Development, Monog. Serial No. 58, No. 1. Lafayette Ind.: Child Development Publications, 1954, **19**, 7–99.
21. EDGINGTON, E. S. Interocular transfer with control for conjugate eye-movement. *Dissertation Abstr.*, 1955, **15**, 1656–1657.
22. FITS, P. M. Use of the eyes in reading instruments. *O-Eye-O*, 1951, **17**, 11–15.
23. GEMELLI, A., COLOMBI, C., & SCHUPFER, R. E. L'enregistrement electrique des mouvements oculogyres et ses applications. *Cont. Lab. Psicol.* Milano: Univer. Sacro Cuore, 1952, Ser. 15, 26–36.
24. GERATHEWOHL, S. J., & STRUGHOLD, H. Motoric responses of the eyes when exposed to light flashes of high intensities and short duration. *J. Aviat. Med.*, 1953, **24**, 200–207.
25. GILBERT, L. C. Functional motor efficiency of the eyes and its relation to reading. *Univ. Calif. Publ. Educ.*, 1953, 11 (3), 159–232.
26. GLOCK, M. D. Effect upon eye movements and reading rate at the college level of three methods of training. *J. educ. Psychol.*, 1949, **40**, 93–106.
27. GRAY, W. S. Summary of reading investigations. *J. educ. Res.*, 1946, **39**, 401–433; 1947, **40**, 401–435; 1948, **41**, 401–435; 1949, **42**, 401–437; 1950, **43**, 401–437; 1951, **44**, 401–441; 1952, **45**, 401–435; 1953, **46**, 401–437; 1954, **47**, 401–439; 1955, **48**, 401–442; 1956, **49**, 401–436; 1957, **50**, 401–441.
28. GRAY, W. S. A study of reading in fourteen languages. In *The teaching of reading and writing: An international Study.* Chicago: Scott Foresman, 1956.
29. HACKMAN, R. B., & TINKER, M. A. Effect of variations in color of print and background upon eye movements. *Amer. J. Optom.*, 1947, **34**, 354–359.
30. HARTRIDGE, H., & THOMSON, L. C. Methods of investigating eye movements. *Brit. J. Ophthal.*, 1948, **32**, 581–591.
31. HENRY, L. K., & LAUER, A. R. A comparison of four methods of increasing the reading speed of college students. *Proc. Iowa Acad. Sci.*, 1939, **46**, 273–276.
32. HILLMAN, H. H. The photographic study of children's eye-movements during reading. *Res. Rev.*, 1955, No. 6, 27–39.
33. HOFFMAN, A. C. Eye-movements during prolonged reading. *J. exp. Psychol.*, 1946, **36**, 95–118.
34. JIMÉNEZ-HERNÁNDEZ, A. Psicología de la lectura: investigación del proceso oculomotor. [The psychology of reading: investigation of the oculomotor process.] *Pedagogia*, 1954, **2** (1), 29–45.
35. KLEITMAN, N., & SCHREIDER, J. E. Diurnal variation in oculomotor performance. *Année psychol.*, 1951, **50**, 202–215.
36. LAYCOCK, F. Significant characteristics of college students with varying flexibility in reading rate. I. Eye-movements in reading prose; II. Motor and perceptual skill in "reading" material whose meaning is unimportant. *J. exp. Educ.*, 1955, **23**, 311–330.
37. LEDBETTER, F. G. Reading reactions for varied types of subject matter: an analytical study of the eye movements of 11th grade pupils. *J. educ. Res.*, 1947, **41**, 102–115.
38. LOFQUIST, L. *Eye movements in a special reading situation.* M.A. thesis, Univer. of Minnesota, 1941.
39. LORD, M. P. Measurement of binocular eye movements of subjects in the sitting position. *Brit. J. Ophthal.*, 1951, **35**, 21–30.
40. LORD, M. P., & WRIGHT, W. D. Eye movements during monocular fixation. *Nature*, 1948, **162**, 25–26.
41. LORD, M. P., & WRIGHT, W. D. Eye movement research. *Light and Lighting*, 1949, **42**, 309–312.
42. LUNDBERG, N. Electric registration of

eye movements. *Acta Otolaryng.* 1941, **29**, 451–455.

43. MANOLAKES, G. The effects of tachistoscopic training in an adult reading program. *J. appl. Psychol.*, 1952, **36**, 410–412.
44. MARG, E. Development of electro-oculography. *A.M.A. Arch. Ophthal.*, 1951, **45**, 161–185.
45. MORSE, W. S. A comparison of the eye-movements of average fifth and seventh grade pupils' reading materials of corresponding difficulty. In *Studies in the Psychology of reading.* University of Michigan Monographs in Education, No. 4. Ann Arbor: Univer. of Michigan Press, 1951, 1–64.
46. MUÑOZ, J. M., ODORIZ, J. B., & TAVAZZA, J. Registro de los movimientos oculares durante la lectura. *Rev. sociedad Argentina Biologia*, 1944, **20**, 280–286.
47. PATERSON, D. G., & TINKER, M. A. The effect of typography upon the perceptual span in reading. *Amer. J. Psychol.*, 1947, **60**, 388–396.
48. PERRY, W. G., & WITLOCK, C. P. A clinical rationale for a reading film. *Harv. educ. Rev.*, 1954, **24**, 6–27.
49. POWSNER, E. R., & LION, K. S. Testing eye muscles. *Electronics*, 1950, **23**, 96–99.
50. RATLIFF, F., & RIGGS, L. A. Involuntary motions of the eye during monocular fixation. *J. exper. Psychol.*, 1950, **40**, 687–701.
51. RIGGS, L. A., RATLIFF, F., CORNSWEET, J. C., & CORNSWEET, T. N. The disappearance of steadily fixated visual test objects. *J. opt. Soc. Amer.*, 1953, **43**, 495–501.
52. SALATINI, R. W. Behavior of vision. *Optom. Wkly.*, 1953, **44**, 1725–1729.
53. SCIPIONE, A. M. Eye movements as related to reading. *Columbia Optom.*, 1953, **27**, 3–4.
54. SIMONSON, E., & BROZEK, J. Effects of illumination level on visual performance and fatigue. *J. opt. Soc. Amer.*, 1948, **38**, 384–397.
55. SIMONSON, E., & BROZEK, J. The effect of spectral quality of light on visual performance and fatigue. *J. opt. Soc. Amer.*, 1948, **38**, 830–840.
56. SOMMERFELD, R. E. An evaluation of the tachistoscope in reading improvement programs. *Third Year-book of the Southwest Reading Conference for Colleges and Universities.* Fort Worth: Texas Christian Univer. Press, 1954, 7–25.
57. SVZ, H. The Lifwynn eye-movement camera. *Science*, 1946, **103**, 628–629.
58. TEN DOESSCHATE, G., & LANSBERG, M. P. Time consumption in eye movements. *Ophthalmologica*, 1954, **128**, 298–300.
59. TILLSON, M. W. Changes in eye-movement pattern. *J. higher Educ.*, 1955, **26**, 442–445; 458.
60. TINKER, M. A. Apparatus for recording eye movements. *Amer. J. Psychol.*, 1931, **43**, 115–118.
61. TINKER, M. A. Eye movements in reading. *J. educ. Res.*, 1936, **30**, 241–277.
62. TINKER, M. A. The study of eye movements in reading. *Psychol. Bull.*, 1946, **43**, 93–120.
63. TINKER, M. A. Time relations for eye-movement measures in reading. *J. educ. Psychol.*, 1947, **38**, 1–10.
64. TINKER, M. A. Fixation pause duration in reading. *J. educ. Res.*, 1951, **44**, 471–479.
65. TINKER, M. A. Perceptual and oculomotor efficiency in reading materials in vertical and horizontal arrangement. *Amer. J. Psychol.*, 1955, **68**, 444–449.
66. TINKER, M. A., & PATERSON, D. G. The effect of typographical variations upon eye movements in reading. *J. educ. Res.*, 1955, **49**, 171–184.
67. TRAXLER, A. E., & TOWNSEND, A. *Another five years of research in reading.* New York: Educational Records Bureau, 1946.
68. TRAXLER, A. E., & TOWNSEND, A. *Eight more years of research in reading.* New York: Educational Records Bureau, 1955.
69. WALTON, H. N. Vision and rapid reading. *Amer. J. Optom.*, 1957, **34**, 73–82.
70. WATERMAN, J. T. Reading patterns in German and English. *The German Quart.*, 1954, **26**, 225–227.
71. WESTHEIMER, G. Mechanisms of saccadic eye movements. *A.M.A. Arch. Ophthal.*, 1954, **52**, 710–723.
72. WESTOVER, F. L. Controlled eye movements versus practice exercises in reading: A comparison of methods of improving reading speed and comprehension. *Teach. Coll. Contr. Educ.*, 1946, No. 917.

Received December 20, 1957.

PSYCHOLOGICAL BULLETIN
VOL. 55, NO. 4, 1958

SEX-ROLE DEVELOPMENT IN A CHANGING CULTURE[1]

DANIEL G. BROWN
United States Air Force Academy

One of the more significant psychosocial developments of contemporary American society would appear to be the relatively fluid state of the sex roles of individuals. Within a single generation, significant changes have taken place in the traditional conceptions of what is masculine and what is feminine. Whether such changes have been abrupt enough to be considered a cultural revolution or sufficiently gradual to be simply degrees of cultural variation is difficult to judge. In either case, however, this changed and changing cultural pattern has a number of implications and possible effects that bear directly on individual, group, and institutional behavior. In this connection such questions as the following might be asked: What are some of these changes that have taken place in the sex roles? Have such changes been more pronounced in the feminine role than in the masculine role? How have these changes affected the life adjustment of individuals? And the relationships of the sexes with each other? What about the effect on boys and girls at the present time and in the years ahead? These are just a few of the problems in the area of masculinity-femininity development and adjustment that need to be studied and investigated.

The present paper is primarily directed toward a consideration of the nature and theoretical implications of sex-role development in children.

DIFFERENTIATION OF SEX AND SEX ROLE

As a starting point, consideration might be given to the age at which the child becomes aware of biological sex differentiation per se as well as when the child becomes aware of the essential meaning of "masculine" and "feminine," i.e., sex-role behavior.[2] At what age for example is the average child able to distinguish between the sexes and to distinguish himself or herself as a boy or girl? Evidence suggests that between two thirds and three fourths of children by the age of three are able to make this basic distinction (**12, 13, 31**).

Evidence also suggests that sex-role differentiation is a gradual process, probably beginning in the second year of life and becoming definitely established by the age of three (**18, 30, 31**). By or during the fifth year most children make a clear differentiation between the more obvious biological cues of maleness and femaleness and psychological cues of masculinity and femininity (**1, 3, 9, 12, 13, 18, 20, 21, 26, 30, 31**). As in the other aspects of psychological development, there are undoubtedly wide individual differences in the clarity with which differences be-

[1] This paper was read at a symposium on "Psychological Implications of Changing Sex Roles," at the annual meeting of the APA, New York, September, 1957. Acknowledgement with thanks is made to Ruth E. Hartley, City College of New York, and to G. D. Ofiesh, United States Air Force Academy, for helpful suggestions concerning this paper.

[2] The concept, *sex role* refers to those psychological characteristics and behavioral patterns that are typical of one sex in contrast to the other sex. The sex role of a person consists of the behavior that is socially defined and expected of that person because of his or her status as a male or female.

tween the sexes are perceived by children.

In any event, whatever the exact age in a particular case, it seems safe to conclude that preschool children as a group become fully aware of the fact that the world is divided into two groups of people and that, depending on whether one belongs to one group or the other, different behavior patterns are expected accordingly. At an early age, then, children are being conditioned to and are actively acquiring their sex roles. One of the most important considerations here has to do with the ***meaning*** and ***significance*** to the child of the earliest perceptions of structural and sex-role differences between boys and girls. What does it mean to a child to become aware of his sex for the first time, and gradually, his sex role? For the child to feel safe, secure, and satisfied in his emerging sexual identity would appear to be one of the most important conditions in his entire development.

Sex-Role Preference in Children

Related to the factor of age in sex and sex-role differentiation in children is the phenomenon of sex-role preference. Does preference for one sex role over the other parallel the developing awareness of the difference between the masculine and feminine roles? Or does preference come later, only after the child has been exposed sufficiently to the differential treatments accorded boys in contrast to girls? The origin and earliest occurrence of sex-role preference is a problem that awaits research investigation. That definite preferences exist in young children for one or the other sex role, however, has been reasonably well demonstrated by several studies (**1, 3, 9, 21, 26**). This problem has been investigated by the present writer by means of a technique known as the *It Scale for Children* (**2**), a scale composed of 36 picture cards, three by four, of objects and figures typically associated with the masculine or feminine roles in our culture (e.g., preferring to play with a tractor rather than a doll; wearing a dress rather than trousers; preferring to be a boy rather than a girl, etc.). A child-figure called "It," relatively ambiguous as to sexual identity, is used in administering the scale by having each child make choices for It, rather than the child himself or herself making the choices directly. Results based on the use of the It Scale with children between the ages of about $3\frac{1}{2}$ and $11\frac{1}{2}$, most of whom were from middle class homes, show that beginning with the youngest preschool group (Ages $3\frac{1}{2}$ to $5\frac{1}{2}$) and extending through the fourth grade (Ages $9\frac{1}{2}$ to $10\frac{1}{2}$) boys express a stronger preference for the masculine role than girls do for the feminine role (**1, 3, 17, 21**). For example, at the kindergarten and third-grade levels, about 85% and 95% of the boys respectively indicate that It would rather be an "Indian Chief" than an "Indian Princess." And when asked which shoes It would rather "dress up and play house in," about 75% and 95% of the kindergarten and third-grade boys respectively chose men's rather than women's shoes.

Girls between the ages of $3\frac{1}{2}$ and $6\frac{1}{2}$ are quite heterogeneous as a group: some are predominantly feminine, choosing practically all of the feminine alternatives; others are predominantly masculine, and still others are "in-betweens," choosing both masculine and feminine alternatives. Taken as a group, for example, 50% express a preference for It

"playing grownups" with cosmetic articles and 50% with shaving articles.

After about the sixth year and extending through the ninth year, most girls show a very strong preference for masculine in contrast to feminine things. For example, between 60% and 70% of the girls in the first, second, third, and fourth grades indicate that It would rather work with "building" tools than with "cooking and baking" utensils.

It is not known whether girls in the fifth grade and beyond (age group from about 10 to 11 and older) become less masculine in preference. Brown's study (**3**) of fifth-grade subjects indicated a definite feminine changeover in girls, but Hogan (**19**) failed to find any such change in the preference patterns of either fifth- or sixth-grade subjects. The whole problem of change in sex-role preference in relation to age needs further and more intensive study.

In contrast to girls, boys *at all ages* show a strong preference for the masculine role. This preference is evident in the youngest group (ages $3\frac{1}{2}$ to $5\frac{1}{2}$) and becomes even stronger until it reaches a near maximum at about the age of eight and thereafter. Thus, between 90% and 95 % of boys in the second, third, fourth, and fifth grades indicate that, given a choice, It would rather wear a shirt and trousers than a dress.

Sex-Role Preference in Children Compared to Adults

To what extent are the sex-role preference patterns of children similar to those of adults? For comparative purposes the Parental Role section of the It Scale may be used (**3**). This section involves asking the child whether It would rather be a mother or a father. Results from this section may be summarized as follows: From about 80% to 95% of boys at all ages from kindergarten through the fifth grade express a preference for It becoming a father, only 5% to 20% for It becoming a mother. On the other hand, in the case of girls from kindergarten through the fourth grade, only about 25% to 45% express a preference for It becoming a mother, while between 55% and 75% for It becoming a father.

These results in the case of children are quite consistent with studies of adults in our culture which asked men and women: "Have you sometimes wished you were of the opposite sex?" or "If you could be born over again, would you rather be a man or a woman?" or "Have you ever wished that you belonged to the opposite sex?" Results may be summarized as follows: only between $2\frac{1}{2}$% and 4% of adult men compared to between 20% and 31% of adult women recall *consciously* having been aware of the desire to be of the opposite sex (**10, 11, 34**). And in Puerto Rico only 33% of a group of adult female students compared to about 93% of male students indicated they would prefer to be female and male respectively if they "could come to life again after death" (**28**). This lopsided preference for being male in preference to being female is also reflected in a recent survey of several hundred university students at Ohio State University who were asked whether they would rather have a male or female child in their family if they could have only one child (**8**). The results showed that 91% of the men and 66% of the women students expressed a preference for a male child. When both groups are combined, boys were preferred by approximately 75% and girls by only 25% of these students.

A significant problem connected with these findings concerns the psychological effect on large numbers of women who openly admit having preferred to be male. How does such awareness affect the self-concept of a girl or woman? The result, according to White (**35**) is to undermine a woman's respect for herself as a woman and to derogate the feminine role in general.

An important anthropological analysis in connection with sex differences in acceptance of appropriate sex roles would be a *cross-cultural* comparison of the percentage of men compared to women who had preferred to be of the opposite sex. Compared to those cultures, for example, where male domination reaches exaggerated proportions, very different results might be expected among the Burmese (**7**), Ojibwa Indians (**33**) and Tchambuli (**23**) where females have relatively high status and a favorable position in their society.

Factors Related to Masculine Role Preference

What factors are functionally related to the much greater preference that boys show for the masculine role than girls show for the feminine role and for the definite preference that many girls show for the masculine role? Although this is a problem in relation to which much research is needed, several conditions or factors may be suggested as contributory.

First, there is the emphasis by Freud on the *anatomical difference* between males and females, the effect of which is supposed to make the boy proud of his status and the girl dissatisfied with hers. Having versus not having a penis allegedly "explains" why girls as well as boys prefer to be boys.

Another attempt to account for sex differences in role preference is the emphasis by Adler on *sociocultural advantages* that go with being male in contrast to being female. The little girl may early perceive the greater prestige and numerous privileges connected with the masculine role. This would tend to arouse envy and drive her in the direction of wanting that which she does not have, namely, masculine status. Adler introduced the concept of "masculine protest" to refer to this phenomenon. That our culture has been and still is masculine-centered and masculine-oriented is obvious. The superior position and privileged status of the male permeates nearly every aspect, minor and major, of our social life. The gadgets and prizes in boxes of breakfast cereal, for example, commonly have a strong masculine rather than feminine appeal.[3] And the most basic social institutions perpetuate this pattern of masculine aggrandizement. Thus, the Judeo-Christian faiths involve worshipping God, a "Father," rather than a "Mother," and Christ, a "Son," rather than a "Daughter."

A third factor relative to the difference between the sexes in role preference is the *greater latitude* of the girls compared to the boys in sex-role development. It appears somewhat paradoxical that, although restricted much more in practically all other respects, girls are allowed *more* freedom than boys in sex-role learning. This is, however, simply consistent with the idea that masculine status is so superior to feminine status that many girls are not even discouraged from striving to attain the former.

[3] Typical examples include: military equipment, cowboy paraphernalia, police badges, airplanes, boats, trains, spaceships, marbles, yo-yoes, miniature auto license plates, etc.

For a girl to be a tomboy does not involve the censure that results when a boy is a sissy. With little, if any, embarrassment or threat, girls may show strong preference for the masculine role; this is not true in the case of boys.

Further evidence of the fact that girls in contrast to boys not only have much more opportunity to pattern their behavior after the model of the opposite sex but in many cases actually do so is cited by Cunningham (6). She reports on a group of fourth- and fifth-grade students who, when asked to describe what they consider to be some of the "pressing problems in human relations" included the following: "How can I stop my sister from being a tomboy?" Other examples that may be cited include:

1. *Clothing.* Girls may wear shirts and trousers with little or no social disapproval, but boys do not wear skirts or dresses; in fact, men who wear feminine clothing, i.e., transvestites, do so at the risk of severe social censure and even legal punishment.

2. *Names.* Many girls are given masculinized names such as Jackie, Stephanie, Billie, Pauline, Jo, Roberta, Frankie, etc., but few boys are given feminized names.

3. *Toys and play activities.* Girls may play with any or all of the toys typically associated with boys (e.g., cars, trucks, erector sets, guns, etc.) but boys are discouraged from playing with toys that are considered feminine (e.g., dolls, dishes, sewing materials, etc.).

Goodenough (14, p. 318) has commented on the greater freedom of girls in sex-typed play as follows: "A boy is not likely to be a Dale Evans, but a girl often becomes Roy Rogers, or any of his masculine colleagues. Boys are rarely glamour girls, but many little girls fall eagerly into the roles of space men, or masculine rough riders."

Based on research findings that show boys consistently making more appropriate sex-typed choices than girls, Rabban (26) and Hurlock (20) conclude that "boys are more aware of sex-appropriate behavior than girls." Rather than being "more aware" than girls, however, it is the relative lack of flexibility of boys in sex-role choices that probably accounts for some of the difference between boys and girls in this regard. Boys simply do not have the same freedom of choice as girls when it comes to sex-typed objects and activities. In this connection, Hartley[4] raises the question as to whether or not results of studies of sex-role preference in children, rather than measuring role preference as such, might not simply reflect the fact that girls are given much and boys little opportunity for variation in expressing preferences for sex-typed objects and activities. This is a good point and should be explored further.

As to the basis of the narrow, rigid sex-typing pattern in males, Goodenough (14) presents evidence that suggests fathers show *greater concern* than mothers for sex-appropriate behavior in their children. In other words, father is more likely than mother to insist that "junior" look and talk and act like a *man*. This pattern, which would tend to have greater impact on the boy than the girl, is consistent with findings presented in the present paper, showing boys are much more likely than girls to make sex-appropriate choices.

Related to these differences in sex roles in childhood appears to be a

[4] Personal communication from Ruth E. Hartley.

parallel difference in adult occupational roles. Even though women traditionally have been subject to various kinds of vocational and economic discrimination, it is still true that a woman may and does enter a "masculine" vocation or profession, e.g., bus driver, engineer, lawyer, etc., with less social disapproval or concern as to one's sex-role "normality" than a man who enters a "feminine" field, e.g., hair stylist, dress designer, nurse, etc. The census in 1950, for example, revealed that women are now in all of the 446 occupations reported by the census. Among the 16,000,000 American women employed, there are "lady" carpenters, sailors, tractor drivers, pilots, telephone linesmen, locomotive engineers, lumbermen, firemen, and even stevedores and longshoremen!

Sex-Role Identification and Sex-Role Preference

In dealing with the complex problem of sex-role behavior it seems particularly important to distinguish between sex-role identification and sex-role preference (**1**). *Identification* is the basic process in which a child, at first involuntarily, and later consciously, learns to think, feel, and act like members of one sex in contrast to the other sex. *Preference* refers to the tendency to adopt the sex role of one sex in contrast to that of the other sex, the former being perceived as more desirable and attractive. With this distinction in mind it is possible to delineate three major sex-role patterns: (*a*) Identification with and preference for the sex role of one's own sex., e.g., a girl may identify with and prefer the feminine role; (*b*) Identification with the sex role of one's own sex but preference for the sex role of the opposite sex, e.g., a girl may identify with the feminine role but prefer the masculine role; (*c*) Identification with the sex role of the opposite sex but preference for the sex role of one's own sex, e.g., a girl may identify with the masculine role but prefer the feminine role. Of the two processes, identification appears to be primary, while preference is more or less secondary relative to sex-role behavior. In normal development the two form a single, integrative process.

In view of the finding that masculine role preference appears to be widespread among girls, it might be hypothesized that conflict or confusion will be conspicuous in their sex-role development. Thus, the fact that girls are destined for feminine functions in adulthood, yet envy and attempt to emulate the masculine role in childhood would tend to produce ambivalence and a lack of clarity in the feminine role (**16**, **24**, **31**). On the basis of a study of sex-role learning in five-year-olds, for example, Fauls and Smith (**9**) refer to the "lack of clear definition" of a sex role in the case of female children. Related to this is the contradiction between the sex-role identification of many girls with the feminine model and the tendency for them to prefer the masculine role.

On the other hand, boys do not necessarily escape difficulties in sex-role development. Even though the culture greatly favors the male, the fact that boys must shift *from* an original identification-attachment with the mother *to* an identification with the father may create difficulties for boys that girls do not experience (**30**). Thus, Sears (**30**) reports that six-year-old boys have not identified with their fathers as well as girls have with their mothers. On the basis of extensive observations of children in

preschools, Hartley arrives at a conclusion similar to that of Sears and, in addition, raises the question as to whether many boys really experience their fathers in their paternal role. She also questions whether many boys even picture themselves as "future fathers" (**18**).

It is also true that a considerable number of boys get overly exposed to the feminine model in early life when the mother is much more prominent in the life of the child than the father. This is especially likely to occur if for any reason the father is psychologically distant or a predominantly negative figure for the son and there is no adequate substitute.

According to Parsons (**25**) and Gorer (**15**) a major effect of the situation in which the father is typically away most of the time while the mother is around continually exemplifying the feminine model is to facilitate the role development of the girl and to complicate the role development of the boy. These writers seem to emphasize the *quantity* of the parent-child relationship rather than the *quality* of such a relationship. In other words, the degree that the child respects, admires and loves the parent may be much more significant than the sheer amount of contact, per se.[5]

Sex-Role Development and Adult Sexual Adjustment

A boy who incorporates the basic features of the feminine model via predominant identification with the mother intrinsically will feel most comfortable in the feminine role, which to him is "normal" and "natural." Such a boy will show a "feminine protest," i.e., he will protest any restriction of his desire and effort to become thoroughly feminine. He will often plead and even demand the freedom to adopt the feminine role (**27**). This is the developmental pattern in childhood that seems to provide the basis for sex-role inversion in adulthood. In fact, inversion refers precisely to the adoption of the basic behavior patterns that are characteristic of the opposite sex (**4***a*).

In cases of males that do not involve a relatively complete inversion of sex role but do show considerable feminine identification, the result may be boys who become rebellious and develop strong defensive reactions in the form of extreme aggressiveness as a means of attempting to counteract their underlying inverted tendencies. MacDonald (**22**) has presented a number of cases of "effeminate" boys who developed pathological aggressive reactions.

Although direct evidence is limited it appears that the child's eventual sexual orientation and adjustment in adolescence and adulthood bears a direct relationship with the nature of his sex-role development in childhood. Adult sexual behavior, at least in part, appears to be an outgrowth of the individual's underlying sex role. Thus, a normal male is one who has identified with, incorporated, and prefers the masculine role; his sexual desire for the female is one aspect of this role. A boy who has identified with, incorporated, and prefers the feminine role will most likely desire a male as a sexual partner in adulthood in keeping with the inverted role pattern. The problem of normal and inverted sex-role development has been discussed in another paper (**4**).

Sex-Role Convergence: A New Cultural Pattern Emerging?

Despite the fact that boys, much more than girls, show a concern for

[5] Acknowledgement is made to L. E. Dameron for making this point in discussion with the writer.

behaving along sex-appropriate lines, there has been considerable change in the direction of both masculine and feminine roles becoming broader, less rigidly defined, less sex-typed, and more overlapping with each other. As Seward (**32**, p. 175) observes, "Today in the post-World War II United States, there is a good deal less self-consciousness about sex roles and probably more freedom of choice for the individual than ever before." In line with this observation is a new course in domestic arts for eighth-graders in a public school in Jersey City, New Jersey, in which boys learn how to cook, sew, and become "efficient housewives," and in which girls learn how to handle "man-sized tools," do woodwork, plumbing repairs, and become the "man-of-the-house." This course is described as so successful that the sexes may be switched in all eighth-grade homemaking and shop courses in the Jersey City system. The same type of course has been established recently in a junior high school in St. Petersburg, Florida. And in the public senior high schools in Denver, Colorado, courses in cooking for boys, metal crafts and lathe work for girls, and child care and training for both boys and girls are offered.

Other indications of the trend toward increasing similarity of sex roles include: (*a*) similarity of educational experiences of girls and boys from kindergarten through the secondary school system; (*b*) husbands doing the dishes, cleaning the house and carrying out other domestic tasks historically considered exclusively "feminine"; (*c*) wives holding down jobs outside the home, many of which have been traditionally "masculine"; and (*d*) the apparel of boys and men that emphasize color, softness, and more delicate features along with the adoption by girls and women of all kinds of "masculine" clothing, hair styles, etc.

Mead (**24**) and Seward (**32**) have pointed out that this greater flexibility in sex-role learning makes for increased interfamily variability and, hence, increasing cultural diversity in this regard. Is it still possible, in our culture for example, to speak of *the* feminine role or *the* masculine role? Or is it necessary to refer to various *roles?* Thus, within a single neighborhood, the role of the husband-father in one home involves almost absolute control, while the role of the wife-mother is strictly subservient and dependent. Next door, the dominating control of the family may be maintained by the wife-mother, while the husband-father is little more than a financially convenient "boarder." Across the street there may be hostile competitiveness and a continual "power struggle" between the husband-father and the wife-mother, each at times emerging "victorious," the other "defeated." And, in still another home, the respective roles of husband-father and wife-mother are largely complementary and equalitarian rather than hierarchical. What must be the effect of these very different parental role patterns on the sex-role identifications and preferences of children who are developing in these respective familial environments? For example, how is the process by which a boy becomes like his father (i.e., "a man") influenced by the various role structures in such families? It is plausible that degree of ease and normality or difficulty and abnormality is directly related to the particular parental role relationships. Intensive study in this area is very much needed.

Finally, on a culture-wide level, the rapid changes in the sex roles of the Japanese during the past decade

might be cited.[6] Among other contributing factors, the cultural diffusion stemming from American occupation of Japan has brought about far-reaching changes, particularly in the feminine role. In a country that gave rise to the expression "as unimportant as a Japanese woman," the traditional and relatively complete subordination of the female to the male appears to be on the way out and is being replaced by a status of women that is beginning to approach that of men. This trend is reflected not only in the fact that women can now vote, an unheard of practice ten years ago, but also in the hopes and aspirations of Japanese children as revealed in their drawings. When asked to draw pictures depicting what they wanted to be when they were grown, many girls drew pictures of teachers, secretaries, industrial workers, beauticians, scientists, etc.

A somewhat parallel development to that in Japan has been taking place in Germany during the past decade or so.[7] Here, too, feminine status has undergone marked change in the direction of greater freedom and opportunity for women in the educational and economic spheres. A continuing sociopsychological analysis of such significant and rapid changes in the feminine sex role of the Japanese and Germans should be very informative and valuable, especially in terms of the impact on the present and future generation of children.

Summary

The young child, as early as the second year of life, begins to distinguish between male and female and between masculine and feminine. Preference for one sex role or the other also begins to emerge early in the life of the child, probably by the third year.

Beginning at the kindergarten level and extending through the fourth grade, boys show a much stronger preference for aspects of the masculine role than girls show for aspects of the feminine role. In fact, a majority of girls in Grades 1 through 4 express greater preference for masculine things than for feminine things. These results are based on the It Scale for Children, a masculinity-femininity projective technique for use with young children.

The finding that girls more than boys show a preference for the role of the opposite sex is paralleled by studies of adults in our culture which reveal that between five and twelve times as many women as men recall having wished they were of the opposite sex.

As to the basis of masculine role preference in both sexes, three factors are mentioned: (*a*) the Freudian emphasis on the anatomical differences between males and females; (*b*) the Adlerian emphasis on sociocultural favoritism of the male compared to the female; and (*c*) the fact that the girl has more latitude than the boy in expressing a preference for sex-typed objects and activities.

A child may identify with and prefer the sex role appropriate to his own sex; or he may identify with and prefer the sex role of the opposite sex; or he may identify with one sex role and prefer the other. A distinction between sex-role identification and sex-role preference is emphasized.

In some ways girls would appear to have a more difficult time than boys in sex-role development; in other ways the development of boys would seem to be more complicated. The general problem of sex differences in

[6] *Life Magazine*, March 29, 1954, **36**, 89–95.
[7] *Life Magazine*, May 10, 1954, **36**, 107–112.

ease of masculinity-femininity development is discussed.

Adult sexual adjustment or maladjustment is related to the nature and outcome of sex-role development in childhood.

There are definite signs that a convergence of the two sex roles gradually is taking place in our society. This cultural trend is evident in the increasing overlap between things and activities formerly considered "exclusively masculine" or "exclusively feminine." A major effect of this emerging cultural pattern is widespread interfamily variability in the sex roles of family members.

Finally, attention is called to the rapid changes in the feminine sex role in Japan and Germany during the past ten years. Emphasis is placed on the need for a continuing sociopsychological analysis of sex-role development in such changing cultures as those of the Japanese and Germans as well as that of our own.

REFERENCES

1. BROWN, D. G. Sex-role preference in young children. *Psychol. Monogr.*, 1956, **70**, No. 14 (Whole No. 421).
2. BROWN, D. G. *The It Scale for Children.* Grand Forks, North Dakota: Psychological Test Specialists, 1956.
3. BROWN, D. G. Masculinity-femininity development in children. *J. consult. Psychol.*, 1957, **21**, 197–202.
4. BROWN, D. G. The development of sex-role inversion and homosexuality. *J. Pediat.*, 1957, **50**, 613–619.

4*a*. BROWN, D. G. Inversion and Homosexuality. *Amer. J. Orthopsychiat.*, 1958, in press.

5. CONN, J. H., & KANNER, L. Children's awareness of sex differences. *J. Child Psychiat.*, 1947, **1**, 3–57.
6. CUNNINGHAM, RUTH, ELZI, ANNA, HALL, J. A., FARRELL, MARIE, & ROBERTS, MADELINE. *Understanding group behavior of boys and girls.* New York: Bureau of Publications, Teachers College, Columbia Univer., 1951.
7. DEIGNAN, H. G. *Burma: gateway to China.* Washington, D. C., Smithsonian Institution, 1943.
8. DINITZ, S., DYNES, R. R., & CLARKE, A. C. Preference for male or female children: traditional or affectional. *Marriage Fam. Living*, 1954, **16**, 128–130.
9. FAULS, LYDIA B., & SMITH, W. D. Sex-role learning of five-year-olds. *J. genet. Psychol.*, 1956, **89**, 105–117.
10. FORTUNE SURVEY, *Fortune*, August, 1946.
11. GALLUP, G. *Gallup poll.* Princeton: Audience Research Inc., June, 1955.
12. GESELL, A., HALVERSON, H. M., THOMSON, H., ILG, F. L., CASTNER, B. M., AMES, L. B., & AMATRUDA, C. S. *The first five years of life: the preschool years.* New York: Harper, 1940.
13. GESELL, A., ILG, FRANCES L., LEARNED, J., & AMES, L. B. *Infant and child in the culture of today: The guidance of development in home and nursery school.* New York: Harper, 1943.
14. GOODENOUGH, EVELYN W. Interest in persons as an aspect of sex difference in the early years. *Genet. Psychol. Monogr.*, 1957, **55**, 287–323.
15. GORER, G. *The American people.* New York: Norton, 1948.
16. GRAY, SUSAN W. Masculinity-femininity in relation to anxiety and social acceptance. *Child Develop.*, 1957, **28**, 203–213.
17. HANDY, G. D. The sex-role preference scale for children: a study of the It-figure. Unpublished master's thesis, Univer. Denver, 1954.
18. HARTLEY, RUTH, FRANK, L. K. & GOLDENSON, R. M. *Understanding children's play.* New York: Columbia Univer. Press, 1952.
19. HOGAN, R. A. Children's sex-role preference with the It-figure. Unpublished master's thesis, Univer. Denver, 1957.
20. HURLOCK, ELIZABETH B. *Developmental psychology.* New York: McGraw-Hill, 1953.
21. LOW, WILLABE P. Sex of the examiner in relation to sex-role preferences in kindergarten children. Unpublished master's thesis, Univer. Denver, 1957.
22. MACDONALD, MARTHA W. Criminally aggressive behavior in passive effeminate boys. *Amer. J. Orthopsychiat.*, 1938, **8**, 70–78.
23. MEAD, MARGARET. *Sex and temperament in three primitive societies.* New York: New American Library, 1935.

24. MEAD, MARGARET. *Male and female.* New York: Morrow, 1949.
25. PARSONS, T. Age and sex in the social structure of the United States. *Amer. sociol. Rev.*, 1942, **7**, 604–616.
26. RABBAN, M. Sex-role identification in young children in two diverse social groups. *Genet. Psychol. Monogr.*, 1950, **42**, 81–158.
27. RONGE, P. H. The "feminine protest." *Amer. J. indiv. Psychol.*, 1956, **12**, 112–115.
28. SANCHEZ-HIDALGO, E. [The feeling of inferiority in the Puerto Rican female.] *Rev. Asoc. Maestros, P. R.*, 1952, **11** (6), 170–171; 193.
29. SCHEINFELD, A. *Women and men.* New York: Harcourt, Brace, 1944.
30. SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing.* Evanston, Illinois: Row, Peterson, 1957.
31. SEWARD, GEORGENE H. *Sex and the social order.* New York: McGraw-Hill, 1946.
32. SEWARD, GEORGENE H. *Psychotherapy and culture conflict.* New York: Ronald, 1956.
33. SHAW, F. J., & ORT, R. S. *Personal adjustment in the American culture.* New York: Harper, 1953.
34. TERMAN, L. M. *Psychological factors in marital happiness.* New York: McGraw-Hill, 1938.
35. WHITE, L., JR. *Educating our daughters.* New York: Harper, 1950.

Received November 26, 1957.

PSYCHOLOGICAL BULLETIN
VOL. 55, NO. 4, 1958

CONTENT AND STYLE IN PERSONALITY ASSESSMENT[1]

DOUGLAS N. JACKSON
Pennsylvania State University
AND SAMUEL MESSICK
Educational Testing Service and Princeton University

In personality theory a ubiquitous and fundamental distinction may be drawn between the interpretation of behavior in terms of (*a*) the *content* of "needs" and of cognitive structures generally and in terms of (*b*) characteristic *styles* of response and action. The separation of these two components of personality organization has taken a variety of forms in the hands of different theorists, as in the Allport-Vernon (**2**) *Studies in Expressive Movements*, in Murphy's (**47**) scholarly discussion of continuity in personality structure, in Klein's (**40**) distinction between needs and control processes, and in Vernon's (**54**) distinction between adaptive and expressive behavior. One may legitimately ask not only *what* a person says or does (the particular content of his statements and actions) but *how* he acts (his characteristic *mode* or *style* of expression).

What is conceptually a relatively sharp distinction is typically blurred and confounded in a particular concrete act; the what and how are fused in a given goal-directed response. An obsequious person indicates his deference not only by the act of yielding, but by the tone of his voice in performing the yielding act. Because content and style are intermixed in a given behavior sequence, and because there is often a theoretical predilection for content components, style is often overlooked in personality assessment. Also, the measurement of content appears to be more direct and unambiguous than the assessment of stylistic dimensions of personality. It is possible, for example, to ask a person what his attitude is on a given topic, or to draw inferences about his need patterns from his reported likes and dislikes (**51**). The obviousness of such devices, while helpful from the viewpoint of labeling what one hopes one is measuring, also permits respondents to distort their scores if they so desire (**32**), something which is less likely to occur in the assessment of style.

In considering the general distinction between content and style, those methods of personality and attitude assessment which are based upon printed questionnaires of one form or another will be emphasized. While the complementary constructs of content and style have special relevance to questionnaire items, where the response-evoking properties of the particular item *form* may contribute markedly to response variance above and beyond the contribution of *content*, the distinction might also be applied usefully to other areas of personality assessment. For example, three possible applications are to perceptual and cognitive style as in the work of Thurstone (**52**), Witkin (**58**), Klein (**39, 40**), Gardner (**21**), and others (**34**); to achievement and apti-

[1] Portions of this paper were read at a symposium on "Experimental Approaches to Personality Assessment" at the American Psychological Association Meetings in New York, 1957.

The authors express their thanks to Lee Sechrest and Riley W. Gardner for commenting on the content and style of the manuscript.

tude testing (**28, 32, 60**); and to the perception of personality (**2, 38, 54, 59**).

The present discussion attempts to do two things: first, to present some evidence showing the important and subtle influences upon responses of stylistic components of item form; and, second, to illustrate how reliable measures of potentially useful stylistic dimensions may be generated from characteristic responses to the form of personality and attitude items as distinct from measures of content.

Personality Style and Response Set

Traditionally, responses to a particular item or set of items are assumed to provide information about the respondent in terms of the item *content*. If, for example, a person agrees with the statement, "Under no conditions is war justified," or answers "true" to the item, "I have more trouble concentrating than others seem to have," it is commonly assumed that these responses, if consistent, will indicate respectively something about the person's attitude toward war or his mental state. Under these conditions response determinants such as the subjects' generalized tendency to agree are legitimately considered as sources of cumulative error, Cronbach's (**13, 14**) familiar "response sets." While Cronbach's emphasis was that response sets often lead to errors of interpretation in the logical validity of tests, he also indicated that these response tendencies might not always be temporary and trivial, but may have a stable and valid component which reflects a consistent individual style or personality trait. While recognizing Cronbach's contribution in describing the phenomenon, it is preferable for the present purposes to change the label from "response set" to components of *style*. This change in terms emphasizes the fact that for certain purposes in personality assessment opportunities for the expression of personal modes for responding should be enhanced and capitalized upon, rather than considered as sources of error to be avoided or minimized. This change also avoids the ambiguity inherent in the concept of "set" (**22**).

Characteristic Styles in Personality and Attitude Questionnaires

Among the more prominent response styles usually evoked by questionnaire items are response acquiescence, overgeneralization, a tendency to respond in a socially desirable way, and the complementary tendencies to respond negativistically, critically, and in a socially undesirable or idiosyncratic manner. Some pertinent illustrations will be drawn of how each of these, operating singly and in combination, may influence the interpretation of responses to psychological tests. Alternative procedures for evaluating these stylistic variables will then be discussed.

Response Acquiescence and Authoritarianism

It has been long recognized that a subject who agrees with a personality or attitude item stated in a positive form may not necessarily disagree with its logical opposite, but may instead show a fairly general tendency toward agreement or disagreement. Studies by Rundquist and Sletto (**49**), by Lorge (**42**), and reviews by Cronbach (**13, 14**), Berg (**8**), and Messick and Jackson (**45**), indicate that response acquiescence is widespread and pervasive over a wide

variety of item content and most pronounced when content is highly ambiguous or imaginary. Berg (**8, 9**) has suggested that acquiescence is a modal response in our culture when the issue before the respondents is unimportant or nonexistent.

The operation of such stylistic tendencies should be taken into account in the course of personality measurement. If a particular *content* area is to be assessed, it is at least necessary to introduce into the scaling procedure appropriate experimental controls for acquiescence, or else reconcile oneself to interpretive equivocality due to the confounding of content and style in a single measure. Other response determinants besides acquiescence, however, must be controlled before characteristics may be unequivocally attributed to respondents on the basis of item content. Messick and Jackson (**45**) have discussed alternative methods for reducing this ambiguity of interpretation in the measurement of authoritarian attitudes.[2]

Even though much of the recent research with the California F scale (**1**) has been of a methodological and critical nature, it nevertheless yields some important information on the relationship between content and style. A number of investigators (**5, 46, 36, 37, 41, 45**) have independently correlated scores based on the California F scale, in which all of the items are so worded that agreement is always scored in the authoritarian direction, with scores based on logically reversed F-scale items. These correlations were not found to be high and negative, as would be expected from consistent responses to item content. With one reversed F scale (**36**), significant positive correlations in the acquiescence rather than the content direction were obtained. Furthermore, there is evidence (**37**) that previously obtained relationships between personality variables and the F scale, formerly thought to be interpretable in terms of correlates of authoritarian ideology or content, may need reinterpretation in terms of consistencies in style. The most recent study requiring such reinterpretation is one by Gilbert and Levinson (**23**), in which a scale purportedly measuring "custodial mental illness ideology" was constructed, with 17 of 20 items requiring agreement to be scored as "custodial ideology." A high correlation between the "custodial ideology" scale and the F scale was used to support the conclusion that "preference for a custodialistic orientation is part of a broader pattern of personal authoritarianism." But Howard and Sommer in a replication[3] found that "custodialism" correlated significantly with agreements to both the original and the Jackson-Messick (**36**) reversed F scales, which would seem to indicate that style rather than content is of primary importance in this instance.

Christie, Havel, and Seidenberg (**12**) have shown that it is possible in some samples to obtain a correlation

[2] Gage, Leavitt, and Stone (**20**) have argued that confounding content and style in the F scale, far from being a source of error, is fortunate, because acquiescence contributes to the empirical validity of the F scale as assessed by independent ratings of authoritarian behavior. If the aim is merely to predict authoritarianism as a criterion, like predicting the success of salesmen, this argument might be legitimate as long as the criterion did not change. But if one hopes to understand the various components of a dynamic construct like authoritarianism, conglomerate indices containing both content and style will not suffice and will confuse the issues (**45**).

[3] Howard, T. W., & Sommer, R. "A Critical Examination of 'Ideology, Personality, and Institutional Policy in the Mental Hospital.'" Unpublished manuscript.

between reversed and original F-scale items significant in the content direction. Jackson, Messick, and Solley (**37**) had previously reported a correlation of +.35 between agreements to original and to reversed F-scale items. What accounts for this apparently considerable discrepancy? One set of investigators predicted and obtained a correlation significant in the acquiescence direction, while another, with a different reversed F scale, predicted and obtained a correlation in the content direction. The answer to this question requires a consideration of more than differences in the content of the two reversed F scales; the form of the items must be examined. Jackson and Messick (**36**) indicated that the original, extremely worded, cliché-ridden style of the F scale was retained in their reversals, while Christie, Havel, and Seidenberg (**12**) explicitly avoided the sweeping generalizations found in the originals and substituted much more cautious statements. It is likely that this difference in item form accounts for the different results of the two sets of investigators. It appears that the tendency to endorse statements containing phrases such as "every person," "no person," "all," "most important," "complete certainty," "never," "must," etc., is a general one, which may act independently of the content. This response style to overgeneralize may contribute to relationships between the F scale and cognitive variables like rigidity (**37**) and perceptual intolerance for ambiguity (**18**). It probably also partially accounts for the frequent observation that verbally elicited ethnic attitudes tend to be highly intercorrelated (**10**), even, for example, in Hartley's (**30**) study where the "groups" were nonexistent and no previous attitude or "cognitive structure" could be assumed to exist. An appraisal of variance associated with aspects of authoritarian content on one hand, and stylistic components like response acquiescence and overgeneralization on the other, would seem to require at least four sets of items: an extremely worded original and reversed F scale and a probabilistic original and reversed F scale. It is suspected that subjects endorsing probabilistic F-scale items would not show as much of the "authoritarian's" intolerance for ambiguity as might be expected, although some relationship between authoritarian ideology and response style might still be obtained.

Response Acquiescence in Personality Inventories

The distinct roles of content and style should also be noted in responses to personality inventories, especially those "true-false" devices like the MMPI developed by the empirical selection of discrimating items. While few, if any, investigators have ever explicitly assumed that the total number of empirically derived scales was the most parsimonious way of summarizing the common variance of an inventory, the use of a large number of separate scales as, for example, in the 9 clinical scales of the MMPI or the 18 scales of Gough's California Psychological Inventory, is justified by the extent to which each makes some independent contribution to the assessment problem not made by the other scales.[4] If there is a great deal

[4] The MMPI was advanced initially as an aid in the prediction of psychiatric diagnoses. In practice it is rarely so used in any literal sense, which is fortunate, as the research evidence (e.g., **7**, **48**) indicates that predictions of specific diagnoses generally cannot be made with certainty. Rather, the original purpose of the MMPI, prediction, has come to be modified so that now scores, singly or in combination, are used to draw inferences about

of common variance among the various scales, this redundancy limits their efficiency.

There is considerable evidence that a very few factors account for the major proportion of the variance on personality inventories of the "true-false" variety. Wheeler, Little, and Lehner (**57**), for example, reported a factor analysis of MMPI scales in which only two major factors and one minor factor were identified. In the light of accumulating evidence it seems likely that the *major common factors in personality inventories of the true-false or agree-disagree type*, such as the MMPI and the California Psychological Inventory, *are interpretable primarily in terms of style rather than specific item content.*

One line of departure from which it is possible to evaluate the role of acquiescence in personality inventories is to consider the percentage of items keyed "true" in each scale as an index of the extent to which that scale elicits response acquiescence. Jackson (**33**) did this with the California Psychological Inventory, computing rank order correlations between the percentage "true" in each scale and the scale's correlation with outside personality measures shown previously to reflect acquiescence. A number of high and significant correlations with such unidirectional scales as the California F scale and the MMPI *K* scale suggests strongly that acquiescence is a major source of variance in the CPI.

Messick and Jackson[5] have obtained evidence of a similar nature for the MMPI. They obtained rank order correlations in the seventies between each scale's percentage "true" and its loading on the first factor as reported in each of several factor analytic studies. Preliminary results suggest that the first factor of the MMPI is interpretable in terms of acquiescence. Equally striking is a recent factor analytic study by Welsh (**55**), who sought to obtain pure-factor MMPI scales through a variant of the internal consistency method. He was rather successful in developing two such scales, labeled A (for anxiety) and R (for repression), which loaded highly on the first and second factors, respectively. The remarkable thing about these scales is that all but one of the 39 items measuring the first factor are keyed "true," while all 40 items for the second pure factor scale are keyed "false." Even though Welsh's two scales are predominantly unidirectional, one in the "true," and the other in the "false" direction, they yield only low negative correlations with each other. This would lead one properly to reject the notion that a simple response set was sufficient to account for all of the variance in the two scales. Nevertheless, each scale does seem to have an acquiescence component, for such a distribution of "true" and "false" items would be unlikely to occur by chance, and Jackson (**33**) has shown that correlations based on both scales correlated significantly with percentage keyed "true" in each CPI scale. Thus, careful consideration must be given to the possibility that response acquiescence is interacting with another variable, either of content or of style, and that responses are determined in part by this interaction. As with the F scale, where acquiescence operates most strikingly in conjunction with statements in the form of

characteristics of respondents (**56**). Somewhat different notions of validity (**15**) and a different mathematical model (**27, 53**) are necessary in the latter case.

[5] Messick, S. J., & Jackson, D. N. "Response Style and Factorial Interpretation of the MMPI." In preparation.

sweeping generalizations, it may be that acquiescence on the MMPI is elicited differentially by certain content categories, or in relation with another stylistic component.

The specific source of the variables which appear to moderate (**50**) the operation of response acquiescence in the MMPI is obviously a complicated research problem which awaits more evidence for a definitive answer. One very promising lead, however, is encountered in another important stylistic determinant of test-taking behavior, the general tendency to endorse socially desirable or socially undesirable statements about oneself. This stylistic response tendency on the part of individuals should be distinguished from the judged characteristics of desirable and undesirable item content. There is considerable evidence that this tendency is general and is related to a tendency to respond in an idiosyncratic or atypical manner. Edwards (**16**) has reported a correlation of .87 between judged social desirability scale values and the proportion of respondents independently endorsing them. Hanley (**29**) obtained correlations of .82 and .89 respectively between probability of endorsement and social desirability ratings for samples of items from the MMPI *D* and *Sc* scales. Fordyce (**17**) correlated with the MMPI clinical scales a set of MMPI items judged to be socially desirable. His obtained correlations were high, ranging from $-.38$ to $-.91$. Although these coefficients indicate the importance of social desirability in scales like the MMPI, they also reflect the influence of response acquiescence, since the social desirability scale contained a disproportionate number of items keyed false. Jackson (**33**) showed that a combination of ranked indices of response acquiescence and social desirability on scales of the California Psychological Inventory was related to the rank of each scale's correlation with the MMPI *K* scale to the extent of $r_s = .86$. This value was higher than the correlation of either response style operating singly, suggesting the possibility of summative effects of response acquiescence and social desirability.

Berg (**8, 9**), granting that there are modal response patterns, suggested that individual differences, particularly deviations, may be revealing of personality style. Berg hypothesized that deviant behavior tends to be general and not specific to any particular content area. Barnes (**3, 4**) appraising the Berg deviation hypothesis in the MMPI, shed important light on the relation between an acquiescent style and idiosyncratic responses. Barnes demonstrated a close correspondence between Wheeler, Little, and Lehner's (**57**) first or "psychotic" factor and total number of items answered deviantly true, and between their second or "neurotic" factor and total number of items answered deviantly false. Although response acquiescence and the tendency to respond in a socially undesirable or deviant manner are confounded in Barnes' analysis, these results strongly support the notion that items judged low in social desirability evoke different tendencies toward acquiescence, as compared with items judged high in social desirability. This interpretation appears consistent with Welsh's (**55**) data, where the first pure factor scale, composed of 38 "true" items out of 39, contains many socially undesirable statements, while the second pure factor scale, where all the items are keyed false, seems to consist predominantly of neutral or somewhat socially desirable statements. Here again, a consistent response style to acquiesce seems to be elicited differentially by

a variety of self-deprecatory statements on the one hand, while, alternatively, neutral or mildly socially desirable statements evoke consistent differential tendencies to disagree or to be negativistic.

Whether there are consistencies attributable to content after allowing for style in these first two factors or, indeed, in any obtained scores on the present form of the MMPI is an important research question, as is the relation between various content and stylistic factors and psychopathology. If Berg (**8**) is correct, if one might just as well use abstract drawings (**3**) as items to discriminate empirically psychiatric patients from normals, then it may be that content is less important and style more important than previously supposed. If this is the case, then past attempts to draw conclusions about respodents on the basis of their answers to uncontrolled item content are suspect. If, on the other hand, consistencies in content can be demonstrated above and beyond components of style, it is extremely important that measures of these content variables make adequate use of proper experimental controls to avoid as far as possible confounding with style. Use of recent advances in scaling theory (**27**, **53**) might be helpful.

Measuring Personality Styles

In approaching the problem of the assessment of style, a curious dilemma presents itself. On the one hand, it is easy to show that most personality tests are loaded with stylistic components, but on the other hand, good measuring devices for these dimensions do not exist, largely because few research workers have attempted explicitly to devise such scales. Typically, a single measure, like the California F scale, the MMPI *K* scale, or Bass's (**6**) collection of aphorisms, has been offered as an index of a response style, acquiescence, for example. Little thought is given to the fact that these measures may not only contain several dimensions of content, but of style as well, thus limiting their usefulness as indices of any particular style. Thus, Fordyce (**17**) has suggested that the MMPI *K* scale reflects tendencies to respond in a socially desirable manner, while Fricke (**19**) has argued that the *K* scale reflects acquiescence. Evidence from each of the two authors is convincing, and, indeed, Jackson's study (**33**) supports the notion that the *K* scale contains both acquiescence and social desirability variance. It may reflect other things as well, but this confounding is not conducive to its use as a measure of one particular style. The same criticism might be leveled at the California F scale, at Edwards' (**16**) social desirability scale, and at Bass's (**6**) social acquiescence scale, all of which seem to confound response acquiescence with social desirability.

One way to construct measures of such styles as acquiescence or overgeneralization would involve selection of items extremely heterogeneous in content. Experimentally independent measures of each style would, of course, be desirable. Since a response style to answer in a socially desirable or undesirable direction seems to be omnipresent, it is hard to avoid in measures of other styles. Rather than attempting to develop items all at one level of social desirability, it might be better to vary social desirability systematically and to observe its relationships and interactions with other variables. Helmstadter (**31**) has described procedures for obtaining separate scores for different components of a test, some of which would be especially relevant to a situation in which one

had already obtained social desirability scale values. Although social desirability has been assumed to be one-dimensional, it is easy to conceive of distinct, but perhaps correlated, dimensions consisting of items reflecting irresponsibility, psychiatric bizarreness, or hostility. The selection of sets of items for different dimensions of judged social desirability would be facilitated by the application of recent advances in multidimensional scaling (**44**). Such refinements as separating out the components of social desirability would do much to clarify response determinants and might put personality evaluation upon a more rigorous basis than has previously been thought possible.

Although the emphasis in this paper has been on some of the more conspicuous stylistic determinants encountered in common personality tests, there are many other possible measures of style that might be derived from personality theory. For example, a tendency to express a liking for diverse things, although it might be response acquiescence in a new disguise, might also represent greater cognitive differentiation or capacity to invest energy freely in objects in one's environment. Such general expressions of "like" and "dislike" have been found to be reliable. On one set of 300 items dealing with diverse activities (**51**), the corrected split-half reliability of the tendency to respond "like" was .86. With a paucity of evidence on these issues, the alternative to such conjecture is carefully planned research, for which there is an obvious need. There are many other research opportunities for the measurement of style, such as asking respondents to select from among two or more personality, attitude, or achievement items, equal in valence or correctness, but couched in different phrasings—perhaps one elaborate and pedantic, one simple, and one containing slang. Preferred modes or styles of expression might also be readily evaluated by techniques disguised as achievement tests (**32**). In this context, it would be interesting to evaluate personality correlates of such attributes as tolerance for logical contradictions within a passage, of a tendency to gamble on achievement tests (**28, 54**), and a variety of other consistent modes of response. Similarly, further research is needed to evaluate Jackson's (**34, 35**) hypothesis that respondents who acquiesce consistently manifest a lower level of cognitive energy in other situations.

Summary

It has been suggested that stylistic determinants, such as acquiescence, overgeneralization, and a tendency to respond in a socially undesirable manner, as distinct from specific content, account for a large proportion of response variance on some personality scales, particularly the California F scale, the MMPI, and the California Psychological Inventory. In developing and evaluating measures of style it is important to select not only those measures which have appeared by accident on already established tests, but to design assessment techniques explicitly to evoke theoretically important styles of response. Research involving response style may contribute to a more systematic measurement in personality and may pay off handsomely in helping to further the common ground between personality theory and personality assessment.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK ELSE, LEVINSON, D. J., & SANFORD R. N. *The authoritarian personality.* New York: Harper, 1950.
2. ALLPORT, G. W., & VERNON, P. E. *Studies in expressive movements.* New York: Macmillan, 1932.
3. BARNES, E. H. Response bias in the MMPI. *J. consult. Psychol.*, 1956, **20**, 371–374.
4. BARNES, E. H. Factors, response bias, and the MMPI. *J. consult. Psychol.*, 1957, **20**, 419–421.
5. BASS, B. M. Authoritariansim or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 611–623.
6. BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.
7. BENTON, A. L. The MMPI in clinical practice. *J. nerv. ment. Disease*, 1945, **102**, 416–420.
8. BERG, I. A. Response bias and personality: the deviation hypothesis. *J. Psychol.*, 1955, **40**, 61–72.
9. BERG, I., & RAPAPORT, G. M. Response bias in an unstructured questionnaire. *J. Psychol.*, 1954, **38**, 475–481.
10. CAMPBELL, D. T., & MCCANDLESS, B. R. Ethnocentrism, xenophobia, and personality. *Hum. Relat.*, 1951, **4**, 185–192.
11. CHAPMAN, L. J., & CAMPBELL, D. T. Response set in the F scale. *J. abnorm. soc. Psychol.*, 1957, **54**, 129–132.
12. CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.
13. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475–494.
14. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
15. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological testing. *Psychol. Bull.*, 1955, **52**, 281–302.
16. EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.
17. FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.
18. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
19. FRICKE, B. G. Response set as a suppressor variable in the OAIS and MMPI. *J. consult. Psychol.*, 1956, **20**, 161–169.
20. GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The psychological meaning of acquiescence set for authoritarianism. *J. abnorm. soc. Psychol.*, 1957, **55**, 98–103.
21. GARDNER, R. W. Cognitive styles in categorizing behavior. *J. Pers.*, 1953, **22**, 214–223.
22. GIBSON, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.
23. GILBERT, DORIS C., & LEVINSON, D. J. Ideology, personality, and institutional policy in the mental hospital. *J. abnorm. soc. Psychol.*, 1956, **53**, 263–271.
24. GOUGH, H. G. Studies of social intolerance: I–IV. *J. soc. Psychol.*, 1951, **33**, 237–271.
25. GOUGH, H. G. Predicting social participation. *J. soc. Psychol.*, 1952, **35**, 227–233.
26. GOUGH, H. G. *California Psychological Inventory Manual.* Palo Alto: Consulting Psychologist Press, 1957.
27. GREEN, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology*, Vol. I. Cambridge: Addison-Wesley, 1954.
28. GUILFORD, J. P., & LACEY, J. I. (Eds.) *Printed classification tests.* Washington: U. S. Government Printing Office, 1947.
29. HANLEY, C. Social desirability and responses to items from three MMPI scales: D, Sc, and K. *J. appl. Psychol.*, 1956, **40**, 324–328.
30. HARTLEY, E. L. *Problems in prejudice.* New York: Kings Crown Press, 1946.
31. HELMSTADTER, G. C. Procedures for obtaining separate set and content components of a test score. *Psychometrika*, 1957, **22**, 381–394.
32. HILLS, J. R. Objective tests of personality for practical use. Princeton, N. J.: Educational Testing Service Research Memorandum 57-4, 1957. (Multilithed report.)
33. JACKSON, D. N. Response acquiescence in the California Psychological Inven-

tory. *Amer. Psychologist*, 1957, **12**, 412–413. (Abstract).

34. JACKSON, D. N. Independence and resistance to perceptual field forces. *J. abnorm. soc. Psychol.*, in press.
35. JACKSON, D. N. Cognitive energy level, response acquiescence, and authoritarianism. *J. soc. Psychol.*, in press.
36. JACKSON, D. N., & MESSICK, S. J. A note on ethnocentrism and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132–134.
37. JACKSON, D. N., MESSICK, S. J., & SOLLEY C. M. How "rigid" is the "authoritarian"? *J. abnorm. soc. Psychol.*, 1957. **54**, 137–140.
38. JACKSON, D. N., MESSICK, S. J., & SOLLEY, C. M. A multidimensional scaling approach to the perception of personality. *J. Psychol.*, 1957, **44**, 311–318.
39. KLEIN, G. S. The Menninger Foundation research on perception and personality. 1947–1952: a review. *Bull. Menninger Clin.*, 1953, **17**, 93–99.
40. KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation*. Lincoln: Univer. Nebraska Press, 1954.
41. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215–221.
42. LORGE, I. Gen-like: Halo or reality? *Psychol. Bull.*, 1937. **34**, 545–546.
43. MEEHL, P. E. The dynamics of "structured" personality tests. *J. clin. Psychol.*, 1945, **1**, 296–303.
44. MESSICK, S. J. Some recent theoretical developments in multidemensional scaling. *Educ. psychol. Measmt*, 1956, **16**, 82–100.
45. MESSICK, S. J., & JACKSON, D. N. The measurement of authoritarian attitudes. *Educ. psychol. Measmt*, in press.
46. MESSICK, S. J., & JACKSON, D. N. Authoritariansim or acquiescence in Bass's data. *J. abnorm. soc. Psychol.*, 1957, **54**, 424–426.
47. MURPHY, G. *Personality*. New York: Harper, 1947.
48. RUBIN, H. Validity of a critical-item scale for schizophrenia on the MMPI. *J. consult. Psychol.*, 1954, **18**, 219–220.
49. RUNDQUIST, E. A., & SLETTO, R. F. *Personality in the depression*. Minneapolis: Univer. of Minnesota Press, 1936.
50. SAUNDERS, D. R. Moderator variables in prediction. *Educ. psychol. Measmt*, 1956, **16**, 209–222.
51. STERN, G. G., STEIN, M. I., & BLOOM, B. S. *Methods in personality assessment*. Glencoe, Ill.: Free Press, 1956.
52. THURSTONE, L. L. *A factorial study of perception*. Chicago: Univer. Chicago Press, 1944.
53. TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, in press.
54. VERNON, P. E. *Personality tests and assessments*. New York: Holt, 1953.
55. WELSH, G. S. Factor dimensions A and R. In G. S. Welsh and W. G. Dahlstrom (Eds.), *Basic readings on the MMPI*. Minneapolis: Univer. of Minn. Press, 1956. Pp. 264–281.
56. WELSH, G. S., & DAHLSTROM, W. G. (Eds.) *Basic readings on the MMPI*. Minneapolis: Univer. of Minn. Press, 1956. Pp. 290–337.
57. WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, **15**, 134–141.
58. WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.
59. WOLFF, W. *Expression of personality*. New York: Harper, 1943.
60. ZIMMERMAN, W. S. The influence of item complexity upon the factor composition of a spatial visualization test. *Educ. psychol. Measmt*, 1954, **14**, 106–119.

Received November 14, 1957.

PSYCHOLOGICAL BULLETIN
VOL. 55, NO. 4, 1958

A NEED FOR ALERTNESS TO MULTIVARIATE EXPERIMENTAL FINDINGS IN INTEGRATIVE SURVEYS

RAYMOND B. CATTELL
Laboratory of Personality Assessment and Group Behavior
University of Illinois

A number of glaring omissions of relevant data have been evident in surveys of particular scientific fields made over the past 10 years. On closer examination it becomes evident that these omissions follow a certain pattern. It seems most desirable to call attention to this unrealized source of distortion in the interests of a fuller use of experimental resources.

In general, the reviewer of a field does a thorough job in terms of collecting, from *Psychological Abstracts*, and the trail leading through bibliographies, the principal findings of univariate experiments, i.e., those using the classical design of varying an "independent variable" and observing the changes in a "dependent variable." But his final list of discovered associations, for the variable or concept on which the survey is focused, frequently shows, by contrast, a shockingly inefficient reporting of the associations found in *multivariate experiments*, e.g., factor analytic experiments, in which his particular concept or variable happens to have been investigated in company with a whole *group* of variables.

It is easy to see how this happens. After setting down the experiments on the subject that are known to him through personal interest, or through discussion with a circle of scientific acquaintances, the person charged with the task of making an integrative survey of the area turns to the literature. As indicated above, he is likely to turn to the index of *Psychological Abstracts* and the annual indexes of a wide array of journals, and to look up all reported studies appearing under the rubric of the given topic, as well as under various possible synonyms for that rubric. If his vision is no wider than that of a training in the traditional, univariate experiment, he then closes the books and feels justified that he has made a thorough search for all relevant studies occurring in the given period of time.

Unfortunately, he has omitted the whole class of research reports in which no reference to all the concepts and variables involved in the given study can be made in the title. It *occasionally* happens in *univariate* studies that the statistically significant finding does not occur in the main relation, but in terms of some "by-product" finding. But it happens systematically and constantly in the area of *multivariate* designs that one study deals with, or reveals, more significant relations, and bears on more concepts, than can possibly be indicated, even with a supreme effort at theoretical condensation, in a title. Most concepts hinge, operationally, upon a variable. Nowadays, factor analytic studies, with the aid of the electronic computer, may work with more than a hundred variables and five thousand relationships, which no title satisfactory to an editor could contain.

Parenthetically, this note is not at all concerned with the relative merits of univariate and multivariate experimental designs. Each has its advantages and disadvantages, its time

for minor and its time for major roles in research. The multivariate design permits a great number of relations of stimuli and responses to be simultaneously examined, but it can examine only linear or roughly linear relations among them. However, as far as surveys of relationships of particular variables are concerned, it is extremely unusual, even in univariate studies, for the integrator to be concerned with statistically significant departures from linear relationships, so there is no argument for omission of multivariate experimental findings on these grounds.

It may be objected that though the *variables* used in a well-conceived multivariate experiment cannot be crammed into the title, such experiments nevertheless deal with relatively few *concepts*, and these should at least enter the title as a guide to the searcher on related themes. Occasionally this can be done, as when a title on, say, primary abilities, indicates most of the theoretical interest attaching to the variables. But in most researches there are as many diverse concepts involved as there are factors, and as many theories as there are relationships among factors. There is, indeed, always a systematically greater number of variables and factors than are directly concerned with the main theme of the study. For a well designed multivariate research recognizes John Stuart Mill's proposition in scientific method—that the definition of "A" depends also on knowing the "not-A" phenomena. Or, to restate this from the statistical standpoint, whereas the univariate experiment is concerned to know what part of the dependent variable variance is associated with the independent variable, and is content to regard the rest as "error," the multivariate experiment secondarily determines *also* where the variance not connected with the main independent variable is tied up.

Consequently the researcher sophisticated in multivariate experimental designs realizes that no matter what the specialized concept—be it dependent or independent variable—on which he is making a survey, it is vitally necessary to peruse the lists of variables and panorama of relationships presented in multivariate researches. This must be done practically without regard to the article titles which the experimenter's predilections or the editor's love of brevity have imposed. The rewards, in avoidance of embarrassing omissions and in discovery of some of the most significant relations, gained from such heightened alertness to data, are very great.

If space permitted it would be regrettably easy to instance surveys in the past 10 years which have been absurdly incomplete through lack of regard for this injunction. For example, there have been surveys on fluency and speed of reaction which have completely omitted the wide range of connections found for such variables in the factor analytic studies of Guilford and his co-workers on creativity. There have been studies dealing with concepts of natural tempo and speed which omit the many associations found in the factor analytic study by Rimoldi. There have been surveys of evidence on the meaning of the galvanic skin response, flicker fusion, rigidity, gestalt closure, color/form reactions, rate of conditioning and extinction, anxiety, persistence of sensory afterimages, reaction to authority, oscillation, suggestibility, response sets in test performance, effects of emotional experience on memory, etc., which have

completely failed to take into account, for their special purposes, systems of significant relationships found between these variables and literally hundreds of others in the present writer's studies on objective personality tests. And there have been surveys of projective principles, of experimental work on defense mechanisms, etc., which have again failed to take into account numerous findings in the factor analytic studies of Eysenck, of the present writer, and of various other researchers who have experimented on these principles and mechanisms in a larger framework of multivariate design.

The irony of this ignorance is that one single multivariate study will sometimes yield more information on the validity or nonvalidity of a set of hypothesized connections than will all the univariate studies which have, up to that point, made up the given bibliography. A single experimental study using a multivariate design with n variables will normally deal with as many relations for a given variable as will n different univariate studies. The surveyor of such studies is not bound to accept the factor analytic solution offered—a fortunate circumstance in view of the small proportion of factor analytic studies reaching acceptable standards of rotational resolution—for he can simply go to the correlations in the original correlation matrix. In the present writer's experience, one glance at such a matrix (on a sufficient sample) has frequently been sufficient to demolish some theory on which a student interested in a particular variable has been about to start a univariate research and to suggest some richer hypothesis with a greater likelihood of survival.

In summary, the call for more systematic attention to the correlation matrices available in multivariate studies, when integrating knowledge on a particular variable or concept, is based on these facts: (*a*) Even one such study usually explores a far greater range of relationships than in two or three dozen univariate studies. (*b*) The form of the data eliminates one of the main difficulties which the survey integrator experiences, namely, that of allowing for effects of slight differences of sample, testing procedure, etc., in combining conclusions on the various relations from many univariate studies. (*c*) In the final inferential reasoning and modification of concepts which are the culmination of such surveys, the integrator of univariate studies is apt to find himself trying to reach a conclusion by now partialling out this influence and now that, or asking what would have happened if this or that had been held constant. Usually this is fraught with much error and conjecture, but in the multivariate study, if he cares to proceed to the factor analysis, this partialling of influences has already been done for him. He is not compelled to go around in a circle of uncertain partiallings out, and even if he should disagree with the particular factor analytic resolution he is free to try a change of rotation. Thus, not only does the omission of a single multivariate factor analytic study from a survey often mean the loss of as much sheer relational information as that of a hundred univariate studies, but it sometimes means also the loss of possible definiteness of conclusions, which could have replaced the speculative adumbrations in which the integrator was finally compelled to take refuge in his attempt to collate many only partially equivalent univariate studies.

Surely one of the main aims of research surveys is to give dependable

glimpses of emerging order in a way that will permit formulation of hypotheses, for further research, that will have a high probability of survival. A sampling of Ph.D. theses in university libraries suggests that graduate students are all too infrequently taught the wisdom of surveying multivariate research findings before choosing their hypotheses. For, indeed, some of the questions they have set themselves will be found already answered in published correlation matrices. But many journal surveys, by psychologists more mature at least chronologically than Ph.D. researchers, unfortunately set no model of acumen in this respect.

Possibly the situation can be remedied without reforming our reading habits simply by some inventor's introducing a mode of labeling and indexing of multivariate studies which is much more subtle than the present ones, using such titles as abilities, anxiety, motivation or temperament, and which will instead give access to the manifold specific experimental variables and relations reported upon in multivariate studies. But miracles are unlikely, and meanwhile our only insurance against the systematic loss of scientific information here described is for compilers of surveys to become more intelligently alert to the numerous relationships of psychological and physiological variables commonly hidden in multivariate studies and to become more conscientiously motivated to seek them out.

Received November 25, 1957.